GL-TR-89-0094

AD-A210 430

A STUDY OF THE IMPACT OF SIMULATED SATELLITE LIDAR WIND OBSERVATIONS

R. N. Hoffman, J.-F. Louis, T. Nehrkorn, M. Mickelson and R. Isaacs

Atmospheric and Environmental Research, Inc.
840 Memorial Drive
Cambridge, MA 02139-3794

30 March 1989

Scientific Report No. 7

Approved for public release; distribution unlimited

GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731-5000

This technical report has been reviewed and is approved for publication.

Donald C. Norquist

DONALD C. NORQUIST
Contract Manager

Donald A. Chisholm

DONALD A. CHISHOLM, Chief
Atmospheric Prediction Branch

FOR THE COMMANDER

Robert A. McClatchey

ROBERT A. McCLATCHEY, Director
Atmospheric Sciences Division

This report has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify AFGL/DAA, Hanscom AFB, MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): GL-TR-89-0094

6a. NAME OF PERFORMING ORGANIZATION: Atmospheric & Environmental Research, Inc.

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Geophysics Laboratory

6c. ADDRESS (City, State, and ZIP Code): 840 Memorial Drive, Cambridge, MA 02139-3794

7b. ADDRESS (City, State, and ZIP Code): Hanscom AFB, MA 01731-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: Geophysics Laboratory

8b. OFFICE SYMBOL (If applicable):

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F19628-86-C-0141

8c. ADDRESS (City, State, and ZIP Code): Hanscom AFB, MA 01731-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
62101F	6670	10	CB

11. TITLE (Include Security Classification): A Study of the Impact of Simulated Satellite Lidar Wind Observations

12. PERSONAL AUTHOR(S): R.N. Hoffman, J.-F. Louis, T. Nehrkorn, M. Mickelson and R.G. Isaacs

13a. TYPE OF REPORT: Scientific Report No. 7

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT (Year, Month, Day): 1989 March 30

15. PAGE COUNT: 86

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Data assimilation; Observing system simulation experiment; Lidar winds; Numerical weather prediction; Satellite observations

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Observing system simulation experiments (OSSEs) are conducted to asses the impact of a space borne Doppler wind lidar sensor. Analyzed and forecast meteorological fields are greatly improved by the addition of the lidar wind profiles. The greatest impacts are in the Southern Hemisphere height and wind fields and the tropical wind fields. In order to calibrate the OSSE results, comparisons are made with real data observing system experiments (OSEs). These comparisons indicate that the expected impact of an actual lidar with the stated geographic coverage (equivalent to that of the current operational temperature sounders) and error characteristics (random, 1 m/s rms) would improve forecasts by approximately 36 hours in the Southern Hemisphere.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Donald Norquist

22b. TELEPHONE (Include Area Code):

22c. OFFICE SYMBOL: GL/LYP

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

1. Introduction

This study presents the results of state of the art observing system simulation experiments (OSSEs) designed to assess the impact of wind profiles observed by a space borne Doppler wind lidar (DWL) (Salvetti, 1987). One of the important objectives of a DWL is improved numerical weather prediction, by improving the initial state specification. These improved analyses would in turn provide an improved data base for diagnosing the global and large scale atmospheric circulation and for understanding climate interactions including the coupling between atmosphere, hydrosphere and biosphere (Curran et al., 1988). To a large degree the success of the DWL may be anticipated and quantified by conducting OSSEs.

As an aid to interpretation our OSSE results are calibrated against results from observing system experiments (OSEs) using real data, described in more detail by Louis et al. (1988). Our study is based on the nature run prepared by the ECMWF and simulated FGGE data base prepared by NMC as described by Dey et al. (1985). There were two previous similar experiments which made use of the same data sets, namely the study by Arnold et al. (1985) at NMC and the study of Atlas et al. (1985) at GLA. In Section 2, to provide background and motivation for the present study we summarize the physical principles underlying lidar wind measurements, the expected instrument characteristics, and the previous aforementioned studies. Section 3 then discusses OSSE methodology in general and the specific experimental design we employed. Results are reported in Sections 4 and 5. Finally, Section 6 contains a summary and our main conclusions.

2. Background

2.1 Principles of Lidar Wind Measurement

Aerosols suspended in the atmosphere can serve as wind tracers for lidar measurements. A photon backscattered by an aerosol particle moving at a wind velocity (v) in the line-of-sight will experience a Doppler shift in frequency of magnitude $(\Delta\nu/\nu) = (2v/c)$, where c is the speed of light. Photons scattered by particles moving toward the observer in the

line-of-sight will experience an increase in frequency, while those scattered by particles receding from the observer will display a decrease in frequency. The Doppler shift caused by aerosol backscatter of a highly stable quasi-monochromatic laser beam could be spectrally analyzed to yield the line-of-sight component of wind velocity. In practice, measuring winds by this method is very difficult because of several facts:

- The Doppler shift is extremely small - a 1 ms^{-1} wind (line-of-sight) results in ~$3x10^{-6}$nm shift at 0.5 μm, or ~$6x10^{-5}$nm at 10 μm.

- The spacecraft velocity, which is about 8 km/sec, also contributes to the Doppler shift. The contribution of this velocity along the line-of-sight of the measurement must be determined to a high accuracy. This places stringent requirements on instruments pointing knowledge and spacecraft attitude.

- The aerosol backscattered signal is often quite weak, particularly in the altitude range of 5-12 km where winds are very important.

- The broadening of the Doppler frequency shift by turbulence can reduce the accuracy of the measurement.

There are two major techniques, using coherent and incoherent detection, to determine wind velocity in this way. The coherent method uses heterodyne detection; mixing the backscattered signal with a local oscillator to yield a beat frequency proportional to the Doppler shift. The incoherent technique measures spectral shifts using a Fabry-Perot interferometer with an array-type detector. The relative advantages and disadvantages of the the two systems, and summaries of potential implementation of DWLs are discussed by Salvetti (1987) and Baker and Curran (1985).

It is now considered technically feasible to measure wind profiles from space by measuring the Doppler shift of a transmitted laser pulse (Salvetti, 1987; Curran et al., 1988). With a strong enough signal the reflected pulse may be range-gated to yield vertical resolution of 1 km or better. The strength of the reflected signal depends principally on the

energy transmitted and the reflectivity of the atmospheric volume being sampled. Since the laser may be focused to a very fine solid angle, horizontal resolution as fine as desired (down to scales of meters) may be obtained. If the reflected signal strengths are sufficient any reasonable desired accuracy might be obtained by this technique.

The lidar measurements are a direct measure of the line of sight (or radial) velocity. Two measurements of the same atmospheric volume from different viewing angles, along with the assumption that vertical velocities are negligible, are required to infer the (u,v) wind components. The simplest method of accomplishing this is to use a conical scan pattern (cf. Fig. 19 of Curran et al., 1988). The assumption of negligible vertical velocity is generally valid. The main exception is cumulus updrafts, which occur inside clouds and, thus, could not be observed.

The global distribution of aerosol is not well known. Since the aerosol concentration directly affects the atmospheric reflectivity to the lidar signal, the relationship of the DWL errors to the transmitted energy cannot be reliably predicted. However, in the experiments described here, we only assume a given error level. The equivalent transmitted energy required might be calculated by assuming some aerosol distribution.

2.2 Instrument Characteristics

From the point of view of numerical weather prediction (NWP), the most important characteristics of any proposed remote sensing system are its geographical coverage, horizontal and vertical resolution and its error characteristics. In a simulation study these characteristics must be accounted for properly. These considerations lead immediately to a number of issues which bear on the interpretation of the results of our experiments.

1) DWL coverage depends on cloudiness. Since the lidar pulse penetrates the cloud only weakly and since multiple scattering from the topmost part of the cloud contaminates any signal received from below, it is impossible to retrieve wind profiles below clouds. In the present experiments, there are no DWL at points below large scale cloudiness.

This is perhaps pessimistic, since on-board shot management might find holes in some of the cloud formations.

2) For NWP, it is not just accuracy of the measurement which is important, the measurement must be representative as well. NWP is really concerned with the spatially and temporally smoothed behavior of the atmosphere. Variations on the scale of meters and seconds, in fact on the scale of kilometers and minutes, are generally considered to be averaged over and are parameterized within the model. Consequently, that part of the measured signal attributable to these scales is considered to be noise from the NWP point of view. This source of error can in some cases be predominant. A prime example is radiosonde observations (RAOBs). When two radiosondes are carried by the same balloon, agreement of the measured quantities is very good, but measurements made by two radiosondes, some distance apart, do not agree as well. The non-representativenes of the measurements is equivalent to an instrumental error. As models improve in resolution, this source of error decreases.

The representativeness issue is of some importance to the current study: We have used simulated data with rms errors of only 1 m/s. Such errors may or may not be attainable for the actual physical measurement, because of limitations on on-board energy and aerosol availability. In any case such error levels are unlikely to include representativeness errors. The DWL naturally averages in the vertical; however, this average is weighted by the vertical profile of aerosol concentration. Our concern lies in the horizontal shot pattern. Since the atmospheric volume sampled may be only meters across, if there is no temporal averaging DWL will contain large errors of representativeness. Averaging many shots would overcome this error source, but shot patterns which have been proposed (e.g. Fig. 20 of Curran et al., 1988) have a typical separation between neighboring shots of 100 km.

3) Spatially correlated errors are difficult for an analysis scheme to remove, because the data tend to corroborate each other. Real data tend to have correlated errors. Even for radiosondes, significant vertical error correlations are present. The data used in the present experiments had data errors which are almost totally uncorrelated. As a result the errors are too easy for the analysis to filter. Atlas (pers. comm.) reports that he undertook some Perfect data experiments in which the random errors were absent. The results of these experiments are nearly identical to the results of the experiments which had random errors.

2.3 Previous Lidar Wind Sounder OSSEs

At a February 1983 Workshop at NMC, NMC, GLA and ECMWF agreed to jointly perform OSSEs to assess the impact of a space borne DWL or WINDSAT (Dey et al., 1985). ECMWF generated the nature run by running their 15 level 1.875 degree resolution global grid point model for 20 days from their operational analysis of 10 November 1979. NMC simulated all FGGE IIb data for the same period by replacing each valid FGGE observation by a simulated observation made up of the value of the nature run evaluated at the observing location plus a simulated observational error. In addition WINDSAT data were simulated at every TIROS sounding location down to cloud top The various procedures used by Dey et al. are recapitulated in Section 3.3 below. Dey et al. envisioned that "[o]nce available, GLAS and ECMWF (as well as anyone else who might be interested) were to conduct a series of fraternal twin experiments with the simulated observations by using their own assimilation/forecast systems." The first two series of these experiments were conducted by Atlas et al. (1985) and Arnold et al. (1985). The experiments described in this report are the third such series of WINDSAT OSSEs.

Atlas et al. (1985) used real data in a control assimilation for 6 days prior to the start of the nature run and then performed 120 h forecasts every fourth day from 00 GMT on Nov. 11, until Nov. 25. They compared Control (or NOSAT), FGGE, Control + TIROS, Control + Cloud Drift Winds (see Fig. 2 in Curran et al., 1988) and Control + Lidar. Control included the conventional data sources (rawinsondes, pilot ballons,

aircraft, surface and ship reports) while FGGE included all data except for WINDSAT profiles. They found significant improvements in the Southern Hemisphere due to adding the WINDSAT to Control. They did not examine the impact of WINDSAT data in the presence of TIROS data however. At a range of 24 to 72 hours they found forecast skill improvements of about 24 hours in the S_1 scores for 500 mb height and sea level pressure for Control + WINDSAT over Control + TIROS.

The analysis system used by Atlas et al. is univariate for the u and v wind components, height and relative humidity and the forecast model is the 4° x 5° 9 layer fourth order grid point model (see Halem et al., 1982, for details). We expect the WINDSAT impact to be understated by the GLA study since the analysis system is a univariate successive correction method and in the extratropics the model will tend to reject the corrections to the wind analysis which are not balanced by corrections to the height analysis. In contrast the experiments reported here make use of a modern multivariate (so-called) optimal interpolation analysis which ensures that the extratropical analysis increments are geostrophically balanced (Norquist, 1988).

In addition Atlas et al. (1985) compared their results for the experiments not making use of the WINDSAT to comparable real data experiments. They concluded that the magnitude of the simulated data impacts were substantially correct in the Southern Hemisphere. However, the Northern Hemisphere skill of the Control experiment was unrealistically high. Presumably this lessens the impact of the additional observing systems in this hemisphere. Furthermore, the impact of TIROS and special FGGE data is overestimated by the OSSEs. Our results agree with these conclusions. To a large extent these differences between OSSEs and OSEs may be due to the simplified nature of the observing errors simulated by Dey et al. (1985).

Arnold et al. (1985) conducted FGGE, FGGE - Rawin wind and FGGE + WINDSAT OSSEs and corresponding FGGE and FGGE - Rawin OSEs. In each case a five day assimilation was followed by a five day forecast. Only lidar winds in the tropics (-30 to +30) were used by the FGGE + WINDSAT assimilation. In the Atlas et al. and in our experiments global WINDSAT coverage

was assumed. Like Atlas et al., Arnold et al. estimated an improvement in forecast skill of approximately one day. However, the greatest impacts in this experiment were seen in the Northern Hemisphere. Their comparison of OSSEs and OSEs indicate that the assimilation system is quite skillful in OSSE mode even with data withheld. As a consequence impacts in the OSSEs of adding data are smaller than in the OSEs (where there is more room for improvement).

Arnold et al. used the 30 wave global spectral model and the multivariate height and wind optimal (or statistical) interpolation on 12 mandatory levels which was operational then (in early 1984) at NMC. At that time the spectral model had relatively simple physics, including only surface exchanges of heat, moisture and momentum over the ocean, surface exchanges of momentum over land, a dry adiabatic adjustment, large scale supersaturation precipitation and a convective Kuo scheme. In both the NMC and AFGL analysis schemes, in the near equatorial belt, the coefficient of geostrophy is smoothly reduced from unity at 25 degrees to zero at 10 degrees. The wind increments, however, are still required to be nondivergent. We remark that utilizing only tropical lidar winds probably significantly limits the impact observed in the Arnold et al. experiments. While tropical winds do exert an influence on the extratropics there was no direct updating of the extratropical heights in these experiments due to the WINDSAT data. In this important respect the Arnold et al. and Atlas et al. experiments are similar. We suggest that the greater impact seen in our experiments is directly due to the updating of the extratropical height field by the DWL observations.

Arnold et al. (1985) describe a very simple calibration procedure. They assumed that the average difference in rms error between different assimilations were linearly related, e.g they assumed the difference in rms errors between the FGGE and FGGE - Rawin OSSEs is proportional to the difference in rms errors between the FGGE and FGGE - Rawin OSEs and similarly for comparisons between FGGE and FGGE + DWL.

3. Experimental Design

OSSE design strategies range from simple insertion of grid point values into unsophisticated models to the use of complex radiative transfer models to simulate data for operational forecast analysis systems. The current study is relatively sophisticated and state of the art. Our experimental design is described in detail in the following sections.

3.1 General OSSE strategy

There are four components common to any OSSE:

1) A four dimensional reference atmosphere, often called the nature run. This is considered to be the "TRUTH".

2) A sampling procedure to obtain simulated observations.

3) A data assimilation system, composed of a forecast model and analysis procedure.

4) A quantitative verification procedure.

Usually, the nature run is simply a long forecast made by an advanced NWP model or Global Circulation Model. The more sophisticated the nature model, the better. Remotely sensed data are influenced by many geophysical parameters, including sea surface temperature, atmospheric aerosol, clouds, etc. In some cases these parameters affect the accuracy of retrievals of other parameters or make such retrievals impossible. These parameters may be responsible for spatially correlated observing errors by inducing local geophysical biases in the retrieved fields (Hoffman, 1988).

The procedures for simulating data from the nature run should be as realistic as possible. The process of simulating the observations should be sophisticated enough to generate realistic observing error statistics. Ideally, errors generated for an OSSE should include the following components:

(1) Representational errors, accounting for the fact that the model does not contain all the scales of motion present in nature.

(2) Geophysical local bias, depending on the sensor type and on the geophysical parameters of the realistic model state. Global biases should be corrected by the data producer and may be ignored in an OSSE.

(3) Random error, which might contain vertical and horizontal correlations. We feel that horizontal correlations are mostly caused by representational errors and geophysical local biases, but vertical errors may be correlated if the sensor retrieval algorithm interrelates several independent observations to a profile of retrieved temperature or other variable.

(4) Sensor filtering. When a sensor uses a statistical retrieval method, all its observations should be filtered by projecting onto the vertical basis functions which are used in the retrieval. This is also true for so-called physical retrieval methods.

The data assimilation system forecast model and analysis procedures should be as realistic and up to date as possible. Impacts of observing systems depend to a certain extent on the forecast and analysis methods used. In some cases, there may be a severe mismatch between the analysis system and the new data. For example, the AFGL analysis system does not presently include a surface pressure analysis; clearly adding a new source of surface pressure data would have no impact. In general it must be noted that operational analysis systems have been tuned for the data they normally receive. The best use of a novel data type may require considerable effort.

Verification of OSSE results is easy because we have total knowledge of the "TRUTH". In these experiments we may legitimately use the word error instead of difference when we compare an experiment to the nature run. Interpretation of these results is not so easy. As noted above, OSSE results are typically too good. The control case is often so good

that there is little room for positive impact. There are two reasons for this: First, the forecast model may be more similar to the model used to generate nature than it is to the real atmosphere. Second, the observational errors are usually too random and easy for the analysis to filter out. For these reasons it is desirable to calibrate the OSSE results to OSE results. In the present case we conduct two OSSEs, NOSAT and STATSAT, for which we have previously conducted analogous OSEs. We will use only a very simple calibration procedure in Section 5. Basically we assume OSSE impacts in statistical measures relative to STATSAT are proportional to corresponding OSE impacts in deriving our estimates of actual WINDSAT impacts.

It is possible to use a series of real analyses for the reference atmosphere, but the results of such experiments would be difficult to interpret for the following reasons. In this situation the "TRUTH" is the actual atmosphere, not the reference atmosphere. Therefore, in data rich areas, the reference atmosphere would agree well with the "TRUTH" while in data voids it would not. Consequently, simulated observations in data rich areas would add correct information, but have little impact because of the concentration of other observations already available, while simulated observations in data poor areas would add erroneous information, which would be carried by the model during the data assimilation cycle to other areas. If the results are then verified in data rich areas we might obtain a negative impact by adding a new observing system. Greater accuracy in the simulated observing system would not avoid adding erroneous data in data poor areas.

3.2 The Nature Run

ECMWF generated the nature run. The nature run is simply a 20 day forecast from the FGGE IIIb analysis produced at ECMWF at 00 GMT 10 November 1979 (Bengtsson et al., 1982). The model used in the nature run forecast was a version of the 15 layer, 1.875 degree grid point model (Hollingsworth et al., 1980). This model included fairly complete physics (Tiedtke et al., 1979) with a diurnal cycle.

To conserve storage space, as we unpacked the gridded nature run tapes, we interpolated the 1.875° grid to a 2.5° grid which we have used for all our data sets and comparisons. A 2.5° grid is substantially finer than the spectral transform grid used by our R30 forecast model and is therefore more than adequate to present our results. In fact the nature run is rather smooth, smoother than many of our analyses and forecasts, and the 2.5° grid is more than fine enough.

3.3 Simulated observations

NMC simulated the FGGE Level IIb and WINDSAT data for the period, in the NMC format (Office Note 29 format) from the ECMWF nature run. Almost all Level IIb data were simulated. However NMC did not simulate constant level balloon data (COBALs), experimental satellite stratospheric sounding data (LIMS) and significant level data. Later GLA converted the NMC data to the standard FGGE format (WMO, 1986). All this work was completed by early 1984. We received copies of the nature run and FGCE format Level IIb data from GLA, courtesy of R. Atlas.

The simulated standard FGGE Level IIb data were created by replacing all the observed atmospheric variables in the real FGGE Level IIb data with values interpolated from the nature run, corrupted by adding a simulated observing error. Therefore if a particular radiosonde report is missing in the real data, it is missing in the simulated data, if it is present in the real data, it is present in the simulated data and has the same quality control marks and missing data flags as the real observation. This yields very realistic data coverage and quality control in the simulated data. However certain discrepancies are possible: For example, cloud drift winds (CDWs) may be present where there are no clouds in the nature run.

The value of the nature run at an observing location is determined by spatially interpolating the nature run at the closest synoptic time (00, 06, 12 or 18 GMT). The vertical interpolation is linear in ln(p) and the horizontal interpolation is quadratic in latitude and longitude for height (Z), wind components (u,v), and temperature (T). For relative humidity (RH), the horizontal interpolations are linear.

The simulated observational error which is added to the value of the nature run at the observing location is composed of a bias and a random Gaussian error which is not correlated with anything else. The size of the random error, or observing error standard deviation (OESD) is appropriate for the particular observation. The OESDs depend on report type, variable and pressure level and are displayed in Table 1 reproduced from Dey et al. (1985). Biases are zero except for TIROS.

The TIROS biases used depend on retrieval path and are displayed in Table 2 reproduced from Dey et al. (1985) which is based on Fig. 2 of Schlatter (1981). In actual practice, the retrieval paths, labeled A, B and C are set depending on whether the retrieval was deemed clear, partly cloudy or cloudy. Note that the OESDs in Table 1 for TIROS also depend on the retrieval path (and are based on Fig. 3 of Schlatter; 1981). In simulation, the retrieval path was determined from the nature run total fractional cloud coverage, f, according to

$$\text{Path} = \begin{cases} \text{A} & \text{if } \quad f \le 60 \\ \text{B} & \text{if } 60 < f \le 90 \\ \text{C} & \text{if } 90 < f \end{cases} .$$

This relationship was tuned to give approximately the same proportion of the different retrieval types as were actually observed on 12 November 1979. The nature run cloud coverage in turn is deduced from the nature run RH as described below. Since the nature run RH field is spatially correlated, the TIROS observational errors will be also. All other errors are uncorrelated.

The simulated WINDSAT data are created at all TIROS reporting locations in a manner similar to that described above for the other data types. At each TIROS location for which NESDIS performed a retrieval, a WINDSAT profile is produced. This profile extends from 10 mb down to the surface in relatively clear conditions or down to cloud top in cloudy conditions. Cloud top is determined as the topmost level for which the fractional cloud in the layer from that level to the top of the atmosphere exceeds 0.9. In the clear cases, winds are retrieved at all mandatory levels: 10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850

Table 1. OESDs Used to Simulate FGGE IIb Data for November

Surface Reports

SFCLT	SFCST	SFCBT	SFCLP	SFCSP	SFCBP	SFCLW	SFCSW	SFCBW
1.0	1.0	1.5	1.0	1.0	2.0	0.0	2.5	0.0

Upper Air Reports

Pressure	1000	890	700	500	400	300	250	200	150	100	70	50	30	20	10
RADT	1.1	1.1	1.1	1.3	1.5	1.7	1.9	2.0	2.1	2.2	2.5	3.0	3.5	4.1	4.5
DRPT	1.1	1.1	1.1	1.3	1.5	1.7	1.9	2.0	2.1	2.2	2.5	3.0	3.5	4.1	4.5
RADW	2.0	2.0	3.0	4.0	5.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
DRPW	2.0	2.0	3.0	4.0	5.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
PIBW	1.8	2.6	3.5	5.0	6.0	8.2	10.0	10.0	10.0	9.2	8.0	6.0	6.0	6.0	6.0
AIRT	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
ASDT	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
AIRW	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
ASDW	2.0	2.0	3.0	4.0	5.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
COBW	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
SATN	4.0	4.0	5.0	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
SATJ	6.0	6.0	6.0	10.0	10.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0	13.0
SATE	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
SATI	4.0	4.0	5.0	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
SATT	4.0	4.0	5.0	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
SATF	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
SATG	7.0	7.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
TIROS Soundings and WINDSAT															
TIRAB	2.0	1.8	1.7	1.8	1.8	1.9	1.9	2.0	2.0	2.2	2.4	2.5	2.6	2.8	2.8
TIRC	3.9	2.5	2.2	2.2	2.0	1.9	1.8	1.7	1.6	1.4	2.5	2.6	2.7	2.8	2.8
WSAT	1.0	1.3	1.6	2.0	2.3	2.5	2.6	2.7	2.8	2.9	2.9	3.0	3.0	3.0	3.0

Legend:

SFCL - Surface land
SFCS - Surface ship
SFCB - Surface buoy
RAD - Radiosonde
DRP - Dropsonde
PIB - PIBAL

AIR - Aircraft
ASD - ASDAR
COB - COBOL
SAT - Satellite cloud tracked winds
WSAT - WINDSAT.

T - temperature (in °K)
W - wind (in ms^{-1})
P - pressure (in mb)
TIRAB - TIROS A&B retrievals
TIRC - TIROS C retrievals

Table 2. Biases used to simulate TIROS temperature retrievals for November

	Retrieval Method		
Pressure Layer	A	B	C
50-70	0.0	0.0	0.0
70-100	-0.7	-0.7	-0.5
100-150	-0.5	+0.15	-0.1
150-200	-0.1	+0.3	+0.4
200-250	+0.5	+0.6	+1.2
250-300	+0.6	+0.5	+0.9
300-400	+0.1	-0.05	-0.15
400-500	-0.4	-0.3	-1.1
500-700	-0.5	-0.35	-1.2
700-850	-0.5	+0.35	-0.6
850-1000	-0.35	+0.3	+1.65

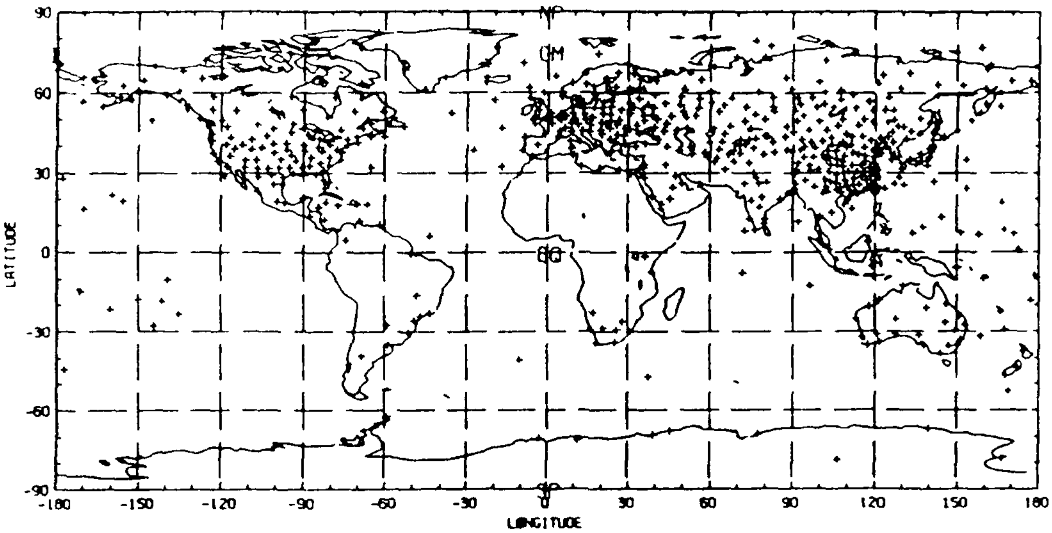
and 1000 mb. In the data base, the WINDSAT data are much more numerous than the TIROS data because we excluded TIROS observations over land. Typically there are 2000 to 4000 WINDSAT profiles per six hour time period. We note that error levels assigned to the lidar winds are approximately half that of the RAOB winds. This characteristic combined with the full TIROS coverage and uncorrelated errors should lead to greatly improved analyses and forecasts.

The nature run cloud fraction at 500, 700 and 850 kPa is determined using a version of Fye's (1978) cloud fraction to RH conversion algorithm. This algorithm is tuned to the ECMWF forecast so that the nature run cloud statistics are reasonable. Any effect of high cirrus cloud on the observation errors is ignored. Layer and total cloud amounts are then calculated assuming random overlap between individual cloud levels within the layer.

Typical data coverage by standard FGGE IIb data is shown in Fig. 1. The data points shown are those actually used by the analysis program, after a gross error and buddy checking quality control; all data with observation times in a 6-hour period centered on the analysis time are shown. The radiosonde locations are those with height observations used for an analysis at 00 GMT on 25 November, at the $\sigma = 0.5$ level. All other plots in Fig. 1 show data locations for 00 GMT on 21 November. Cloud track winds are most numerous at low levels ($\sigma = .86$ and .72) and upper levels ($\sigma = .27$ and .22), whereas aircraft wind reports are concentrated at upper levels ($\sigma = .27$ and .22). Coverage by the TOVS data, shown here for height observation at the $\sigma = 0.5$ level, illustrates the satellite tracks and the fact that retrievals over land were not used. Fig. 2 shows the WINDSAT data at 21 November at 0000 GMT. Areas where the satellite swath is not completely filled with WINDSAT data is a result of clouds.

The simulated data described here are quite complete and realistic, yet have two major failings. First the observational errors are uncorrelated. For example, consider the CDWs: The typical CDW OESD used (8 m/s) is of the proper size, but real CDW errors have large horizontal correlations due to height assignment errors which are responsible for the largest part of the OESD. In reality, then, the CDW errors are not much reduced by the filtering of the analysis procedure. In the simulation

a



b

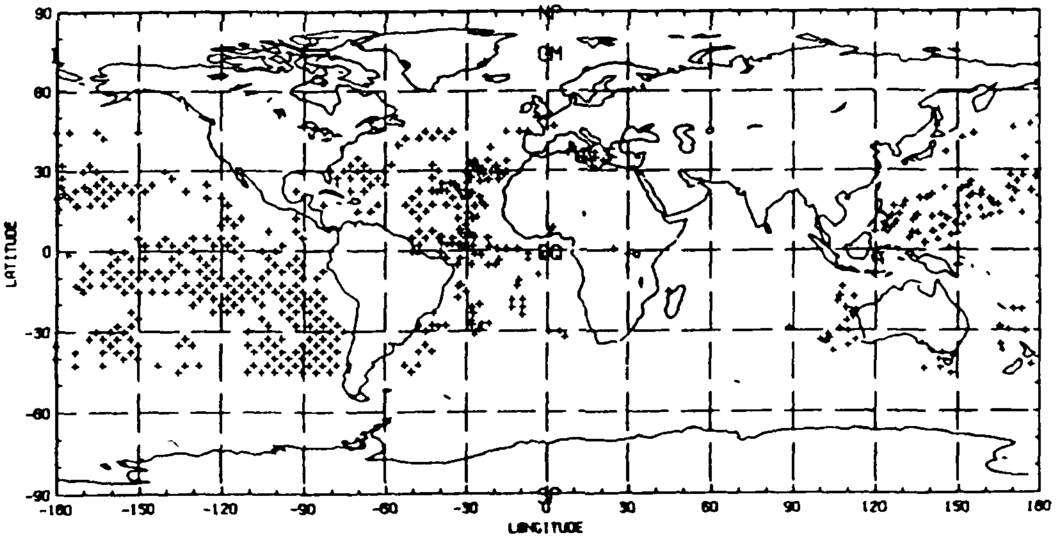
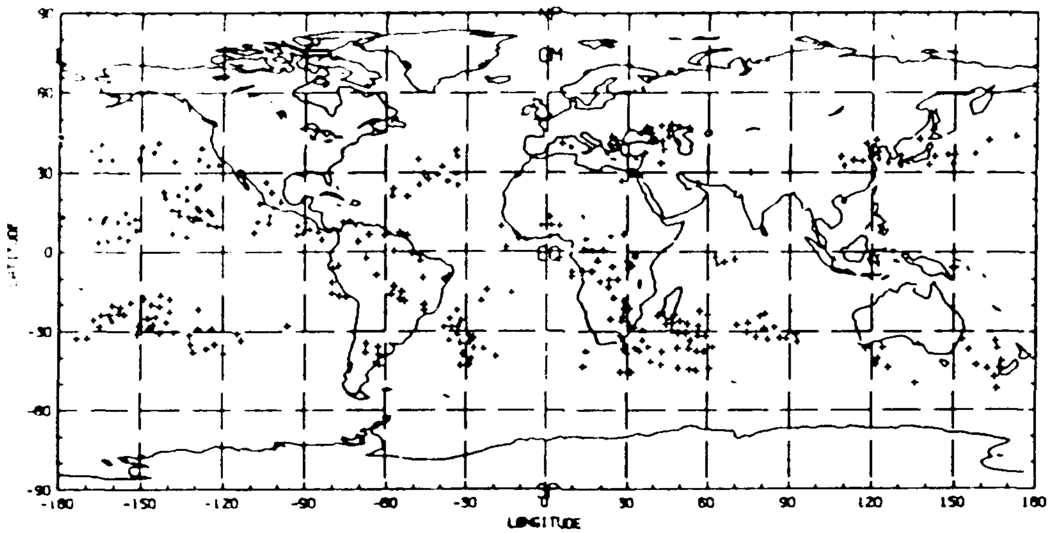


Fig. 1 Data coverage for standard FGGE IIb observations. RAOB heights at σ=0.5 (a), cloud track winds at σ=0.86 and 0.72 (b), and σ=0.27 and 0.22 (c), aircraft winds at σ=0.27 and 0.22 (d), and TOVS heights at σ=0.5 (e).

c



d

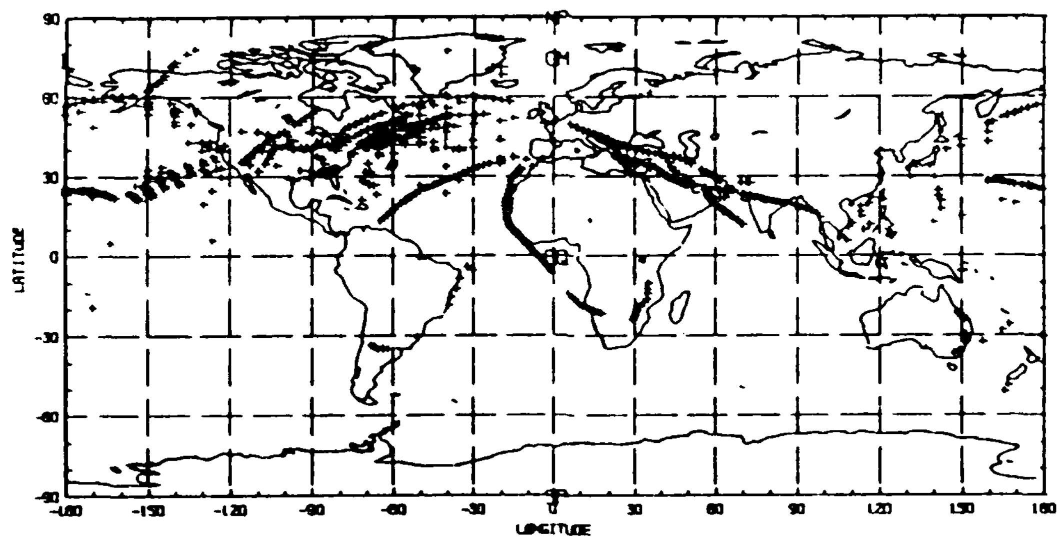


Fig. 1 (continued)

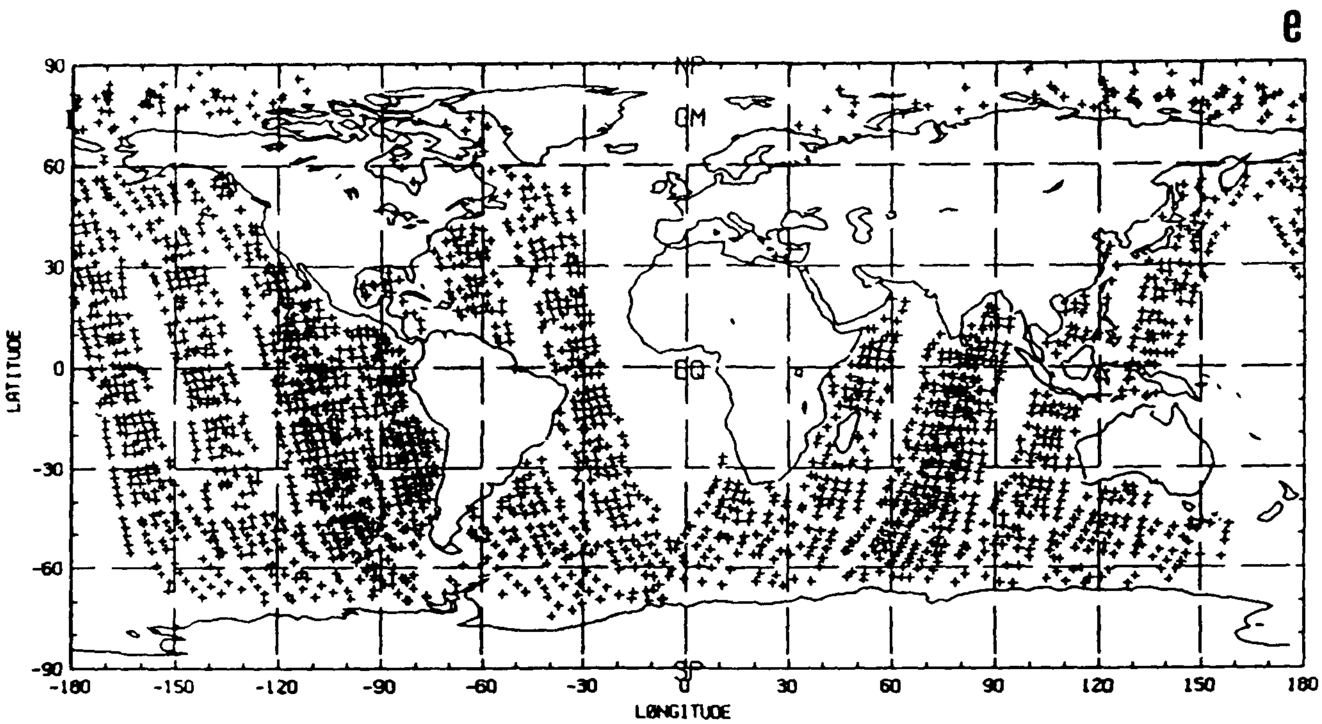


Fig. 1 (continued)

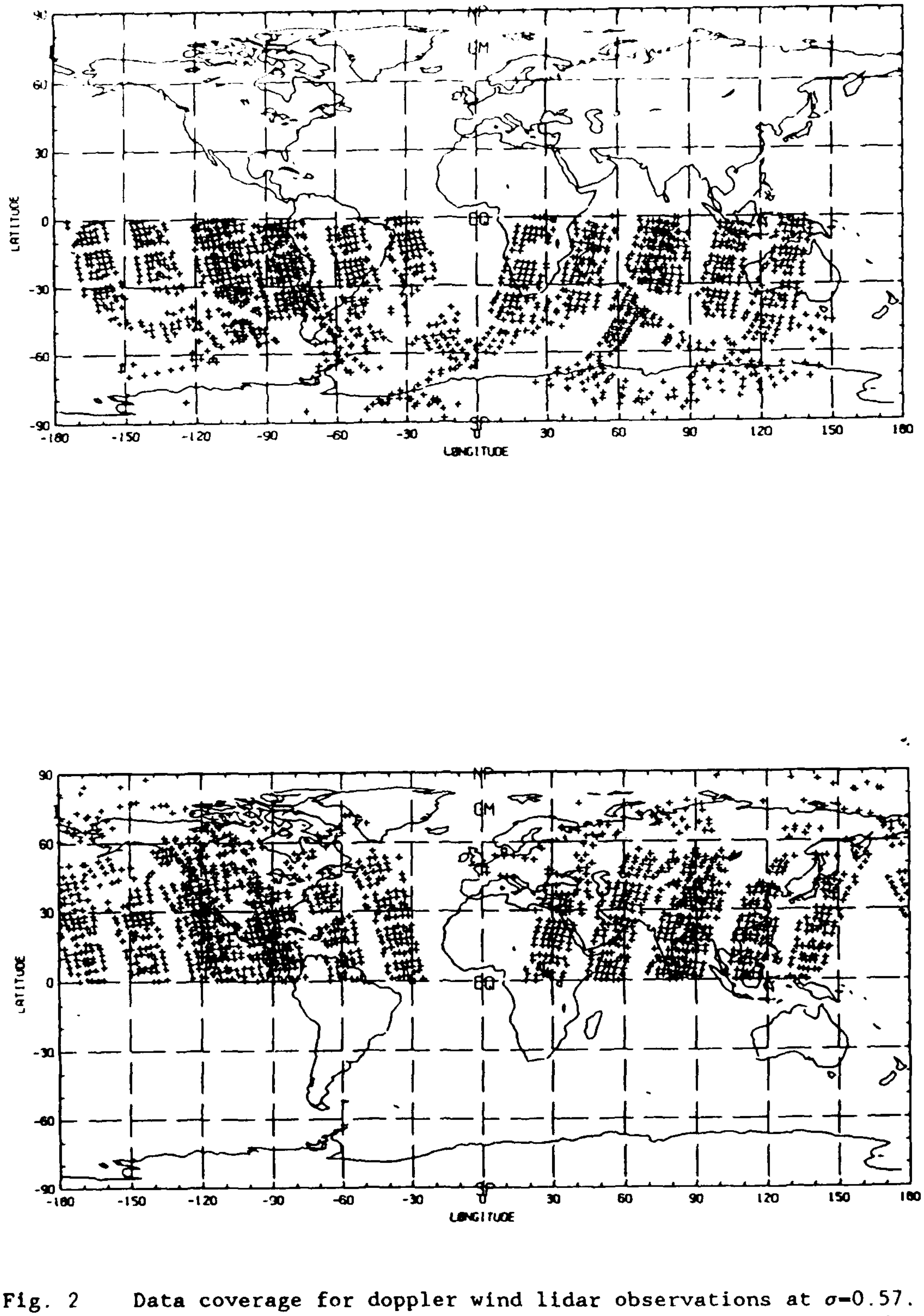


Fig. 2 Data coverage for doppler wind lidar observations at σ=0.57. (There are two separate plots because the data were stored as two separate files.)

experiments described here, on the other hand, the analysis is able to average out the CDW errors effectively because they are uncorrelated and the observations are dense. Second the nature run has little energy in the smallest scales. Small scale energy present in the real atmosphere must be considered part of the observational error and thereby induces spatially correlated errors for all observations. The absence of this source of error also contributes to the (unrealistic) ease with which the analysis averages out the observational errors in the simulation experiments. It would be desirable to "unfilter" the nature run to recreate more realistic small scales, as suggested by Hoffman (1988), but this procedure was not followed here.

3.4 AFGL Forecast and Analysis System

Each simulated data assimilation experiment (OSSE) described here consists of one assimilation run for seven days and three forecasts, each four days in length. The real data assimilation experiments (OSE), used for comparison, consist of two of these assimilation runs. Each assimilation run consists of a series of assimilation cycles, and each cycle in turn is made up of a 6-hour forecast that serves as a first guess of the analysis, an optimum interpolation analysis which combines the first guess fields with the observations, and a nonlinear normal mode initialization of the analysis. The initialized analysis is the starting point for the next 6-hour forecast, which is then used as the first guess of the subsequent assimilation cycle. The forecast model used for the 6-hour forecast is a complete global spectral model (GSM). This model is also used to produce forecasts out to 4 days starting from days 3,5, and 7 of the assimilation run.

3.4.1 Analysis

The AFGL Statistical Analysis Program (ASAP) (Norquist, 1986, 1988) was developed from the NMC multivariate optimal interpolation (OI) procedure as described by Bergman (1979) and by McPherson et al. (1979). The ASAP OI is a multivariate analysis of height and wind components and a univariate analysis of relative humidity, both in model sigma layers. The corrections for an analysis grid point are weighted sums of surrounding

observation-minus-first-guess residuals. The equations for these weights as well as the computation of the horizontal and vertical correlation functions equatorward of 70° latitude is included as described by Dey and Morone (1985) without changing the Bergman formulation (including map factor) for latitudes poleward of 70°latitude. The analysis takes place in the sigma coordinates of the model on a Gaussian grid of 62 x 61 latitude-longitude points.

Data used by the height-wind analysis include Type 1 observations (radiosondes, pibals, etc.), Type 2 observations (aircraft), Type 4 observations (cloud drift winds (CDWs)). The Type 3 surface observations are not used at all. This implies that satellite "heights" are anchored only by the 6 h forecast in regions where radiosondes are absent. The moisture analysis used in the experiments reported here uses only Type 1 data. In all experiments the CDW data were combined (i.e. locally averaged) into "super-obs". There are two principal reasons for doing this: First, to limit the total number of observations, so that computer memory restrictions are not exceeded, and second, the CDW errors are strongly correlated horizontally because the main error source is due to height assignments. Satellite temperature profiles are not used over land in any of the experiments, but all the lidar wind profiles are used.

The data selection algorithm was altered to recognize the WINDSAT data. The basic data selection algorithm follows Bergman (1979) as described by Norquist (1988). The first stage of data selection is done in terms of profiles: Data items in the up to 8 closest profiles are candidates to be chosen for use in the actual analysis in the second stage. Here closeness is measured by the magnitude of height-height forecast error correlation times the number of non missing data items in the profile times a measure of data quality. (A Z observation is considered one data item and a (u,v) pair is considered one data item.) The data quality is taken to be unity for all data except it is 0.42 for satellite temperatures and WINDSAT data. Thus RAOB and other type 1 data are preferred over all others. When a WINDSAT profile and a TIROS profile colocate we store all these data together as one profile. Such profiles are preferred over WINDSAT only profiles which are in turn preferred over single level aircraft and CDW data. Actually, it might have been better to prefer DWL

data to RAOB data, as is suggested by comparing the Northern and Southern Hemisphere results below. The second stage of data selection, which was not altered, selects for each analysis point up to 10 data items from the selected profiles which individually would give the largest reduction in estimated analysis error. In this second stage the actual estimated OESDs are used so that WINDSAT winds would be preferred to RAOB winds.

Since the statistical models used by the OI are never exact in practice, we decided for convenience and realism to leave most of the statistical models in the OI as they were for the real data OSEs described by Louis et al. (1988). The statistical models and parameters used are identical to those described by Norquist (1986), which in turn are based on NMC practice as described by Dey and Morone (1985). Not surprisingly, the OESDs in Table 1 generally agree with the values used by Louis et al. (1988). In particular, the OI assumes RAOB observational errors are correlated vertically and that satellite height observational errors are correlated vertically and horizontally. A number of studies colocating satellite and radiosonde height data were performed by Louis et al. (1988) and slightly different models and parameters for the satellite height observational errors were used in the OSEs reported here. These observational errors are described in detail in Section 2.3 of Louis et al. (1988). In the OSSEs, we use the WINDSAT OESDs directly from Table 1 and a new version of the RAOB height OESDs. The RAOB height OESDs were calculated from the corresponding temperature OESDs listed in Table 1, assuming no vertical correlations of the temperature errors and assuming a 1000 mb height OESD of 10 m, uncorrelated with the temperature errors in the layers above. This yields Z OESDs of 45, 45, 45, 44.7, 44.5, 36.5, 27.5, 25, 25, 20, 12.5 and 10 m for the mandatory levels from 50 to 1000 mb.

3.4.2 Forecast and Initialization

The AFGL normal mode initialization (NMI) is based on the NMC NMI (Ballish, 1980). The AFGL global spectral model is based on the NMC GSM designed by Sela (1980). For the version used here, the physics routines are taken almost intact from NMC (circa 1983). The hydrodynamics, i.e., the adiabatic, inviscid dynamics including vertical and horizontal

advection, time stepping, and transformations between spectral and physical space, were completely redesigned, as documented by Brenner et al. (1982, 1984).

There are a number of parameters in the forecast and initialization codes that can be adjusted. Briefly, the spectral resolution of the forecast model is defined by a rhomboidal truncation at wave number 30. The Gaussian grid of the forecast model contains 76 x 96 latitude longitude points. There are 12 layers, the first (top) 5 of which have no moisture. The sigma interfaces are at 0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.375, 0.50, 0.65, 0.80, 0.925 and 1.00. The time scheme used is centered semi-implicit with a time step of 17.25 minutes. Horizontal fourth order diffusion ($-\kappa\nabla^4$) is applied to all modes of divergence and to modes in the upper half of the rhomboid for vorticity, temperature and specific humidity. The diffusion coefficient used is $\kappa = 6.10^{15}\ m^4\ s^{-1}$. In the NMI, two Machenauer iterations are applied to modes for the four largest equivalent depths which have periods less than or equal to 48 hours.

3.5 Verification Procedures

Our verification procedures include subjective comparisons of analyses and forecasts, quantitative comparisons of rms errors of analyzed and forecast fields and a calibration of the quantitative measures making use of the OSE results. These procedures will be described in more detail in the next section. Here we describe some diagnostic fields which we use.

For verifying the analysis we used the actual nature run as a reference. In this case the calculated difference statistics are in fact error statistics. For verifying the forecasts we made use of the Level II observations. Such a comparison is more realistic for calibrating with the results of our OSE results.

Since the WINDSAT data are direct wind measurements, we anticipated a substantial impact on the wind field. In particular one might hope for improved small scale analyses and improved analyses of divergence. However, the analysis procedure does not allow for divergent wind increments, so that the improvements to the divergence field can only be obtained by

induction as a second order effect. Since the divergence field is so noisy, we used a spectral truncation of triangular 15 before plotting the results.

In measuring the impact of the DWL data another performance measure we use is the rms and bias error of a diagnosed total cloud fraction. Although the "real" simulated clouds generated during the nature forecasts were not saved, the large scale (i.e., nonconvective) cloud fraction is readily diagnosed from the RH field. This sort of measure puts a premium on proper forecasting at the higher range of RH, since errors at lower values have no effect on large scale cloud amounts.

Cloud cover at each of the six moisture carrying mandatory levels was inferred by inverting the Tibaldi formulation as given by Norquist (1988, Appendix). Low, middle and high cloud were then formed from the level cloud by assuming random overlap and grouping pairs of mandatory levels. For example low cloud is calculated as the maximum of 1000 mb and 850 mb cloud. Total cloud was then calculated assuming random overlap between low, middle and high cloud. That is

$$(1 - f_{total}) = (1 - f_{low})(1 - f_{middle})(1 - f_{high}) \quad .$$

Since the relative humidity field and the cloud cover field are dominated by small scales, we have applied a T15 spectral truncation before plotting global maps of these fields.

3.6 SPINUP Experiment

To begin our experiments we first performed two 96 h forecasts starting from "perfect" initial conditions on 00 GMT 11 and 21 November 1979. These initial conditions are taken directly from the nature run, interpolated to the model sigma structure, analyzed into spherical harmonic coefficients and initialized with the adiabatic NMI. We found the growth of forecast error (i.e., forecast - nature) to be rather slow in these forecasts. In fact the forecast ending on the 00 GMT 15 November was not much worse than a typical analysis. Accordingly, this state was

used as the initial "analysis" for a three day SPINUP assimilation experiment, ending 00 GMT 18 November. SPINUP makes use of the standard STATSAT configuration.

3.7 Schedule of Impact Experiments

The end of the SPINUP assimilation is used as the starting analysis for all our OSSEs which therefore run from 00 GMT 18 November through 00 GMT 25 November. (The first analysis of each OSSE is at 06 GMT 18 November.) For each OSSE 96 h forecasts are made from 00 GMT 21 and 25 and in some cases 23 November.

The OSSEs described here are WINDSAT, STATSAT and NOSAT. STATSAT includes all the Level II data excluding WINDSAT which were simulated by NMC as described in Section 3.3 except that surface observations are not used and satellite temperature soundings over land are not used. In NOSAT the satellite temperature soundings and CDW observations are excluded, while in WINDSAT, the doppler lidar wind observations are added.

The OSEs STATSAT and NOSAT are analogous, at least as far as data usage is concerned. The OSEs, however, were run for one week each during February and June 1979. These experiments are described in detail by Louis et al. (1988) and are used here primarily to calibrate the OSSE results. One notable difference between the OSSE and OSE experiments is that the OESD for satellite heights for STATSAT and NOSAT in the February OSE were substantially larger than in the other experiments. (These values are given by Louis et al., 1988.)

4. OSSE Results

4.1 Subjective Synoptic Evaluation

The spinup forecast, starting from perfect initial conditions at 00 GMT 11 November, and the spinup assimilation, which was performed from the end of the spinup forecast at 00 GMT 18 November were compared with the corresponding nature data. The spinup forecast had very similar

500 mb height patterns, but our forecast is consistently warmer than the ECMWF forecast. The possible reason for this warm bias is the lack of a radiation parameterization in our model: a typical radiative cooling of the atmosphere of 1 K/day would correspond to a 500 mb height difference of roughly 20 m/day, which is consistent with the approximate height difference of 80 m at the end of the 4-day forecast. The analyses during the spinup assimilation, which correspond to a STATSAT configuration, are also quite similar to the nature data, except that they are considerably noisier, possibly due to the warm bias of the first guess, which is only corrected at data locations.

The nature run 500 mb height field is shown in Fig. 3 for 00 GMT 23 November. The 500 mb height pattern of the nature data shows a distinct wavenumber four pattern, which is present throughout the entire November time period. Several smaller scale, mobile troughs are superimposed on the long-wave structure. The analyses from the OSSEs shown in Fig. 4 are visibly noisier than nature.

WINDSAT errors appear to be smaller by roughly a factor of two in some areas. (Note that the WINDSAT error plot uses a contour interval of 2 dm while the STATSAT error plot uses a contour interval of 4 dm.) WINDSAT tends to yield higher heights than the reference in the polar region while STATSAT tends to yield lower heights.

Fig. 5 shows the 48 forecast results valid at the same time as Figs. 3 and 4. The forecast errors are much smaller for WINDSAT. Clearly the WINDSAT forecasts are superior. Later, in Section 5, we will analyze differences between the forecasts and the radiosondes. We note here that the largest errors are concentrated towards the pole, while the radiosondes are concentrated in middle latitudes. Examination of a sequence of maps suggests that in the Northern Hemisphere WINDSAT is better by about 24 hours relative to STATSAT.

In Fig. 5, we see the forecast error patterns for STATSAT and WINDSAT are similar although, the magnitudes are different. On the other hand in Fig. 4 the analysis error patterns were markedly different. This suggests that model errors are significant in these experiments.

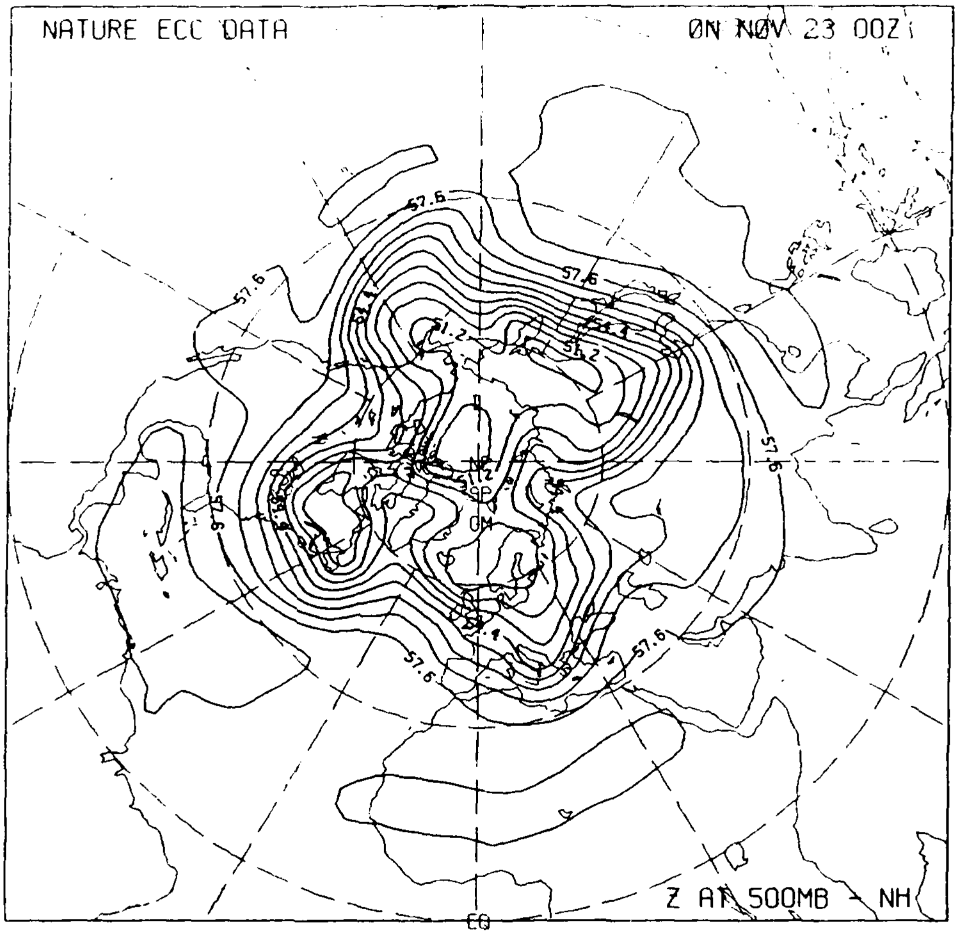


Fig. 3 The nature run 500 mb Northern Hemisphere height field at 00 GMT 23 November 1979 in hundreds of meters. Here and below the 500 mb height field is displayed with an 80 m contour level.

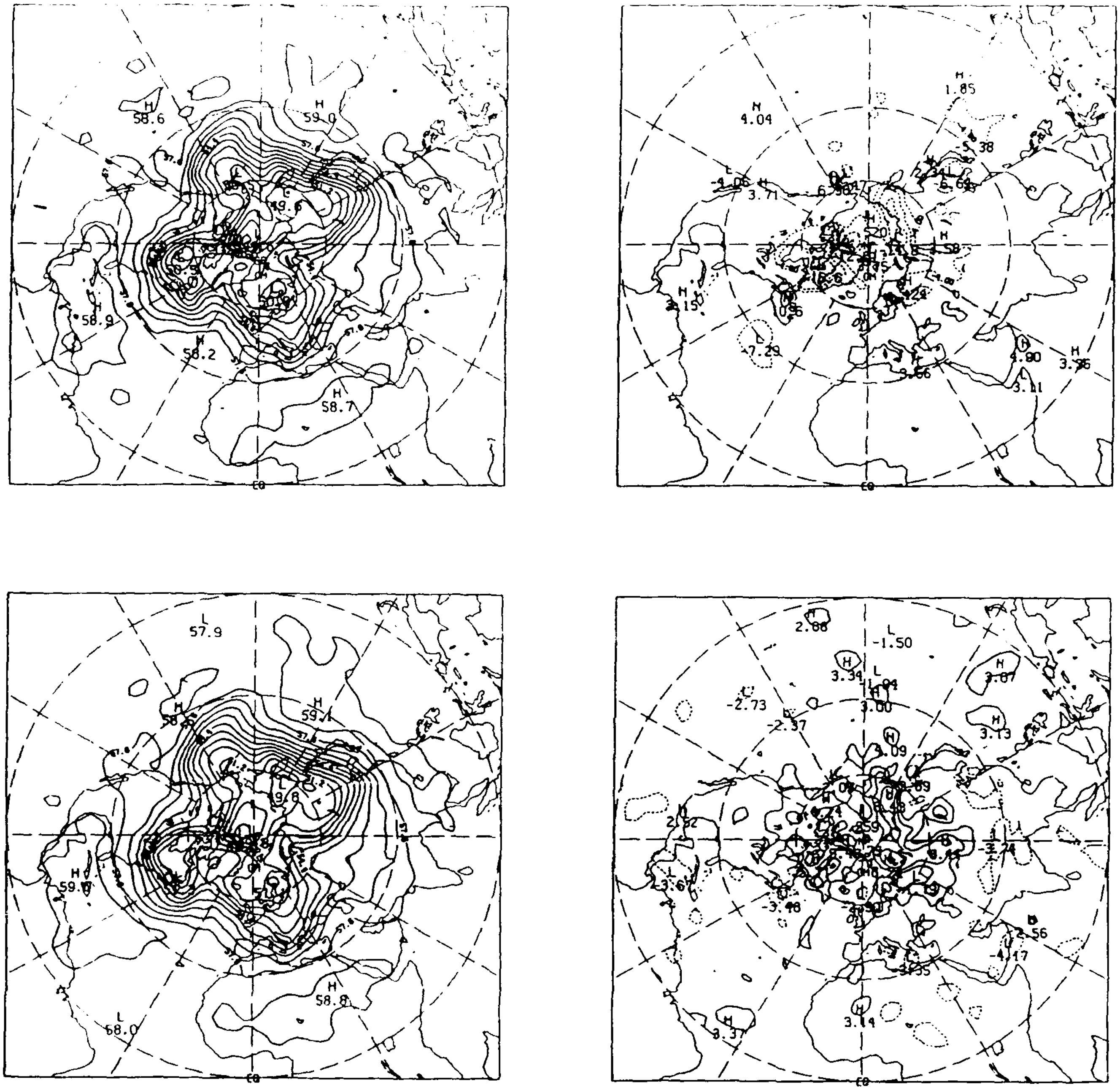


Fig. 4 The STATSAT (top) and WINDSAT (bottom) 500 mb Northern Hemisphere height field analyses (left) in hundreds of meters and analysis error (right) in tens of meters at 00 GMT 23 November 1979. The 500 mb height errors are displayed with a 40 m contour interval for STATSAT and a 20 m contour interval for WINDSAT. Negative values are dashed.

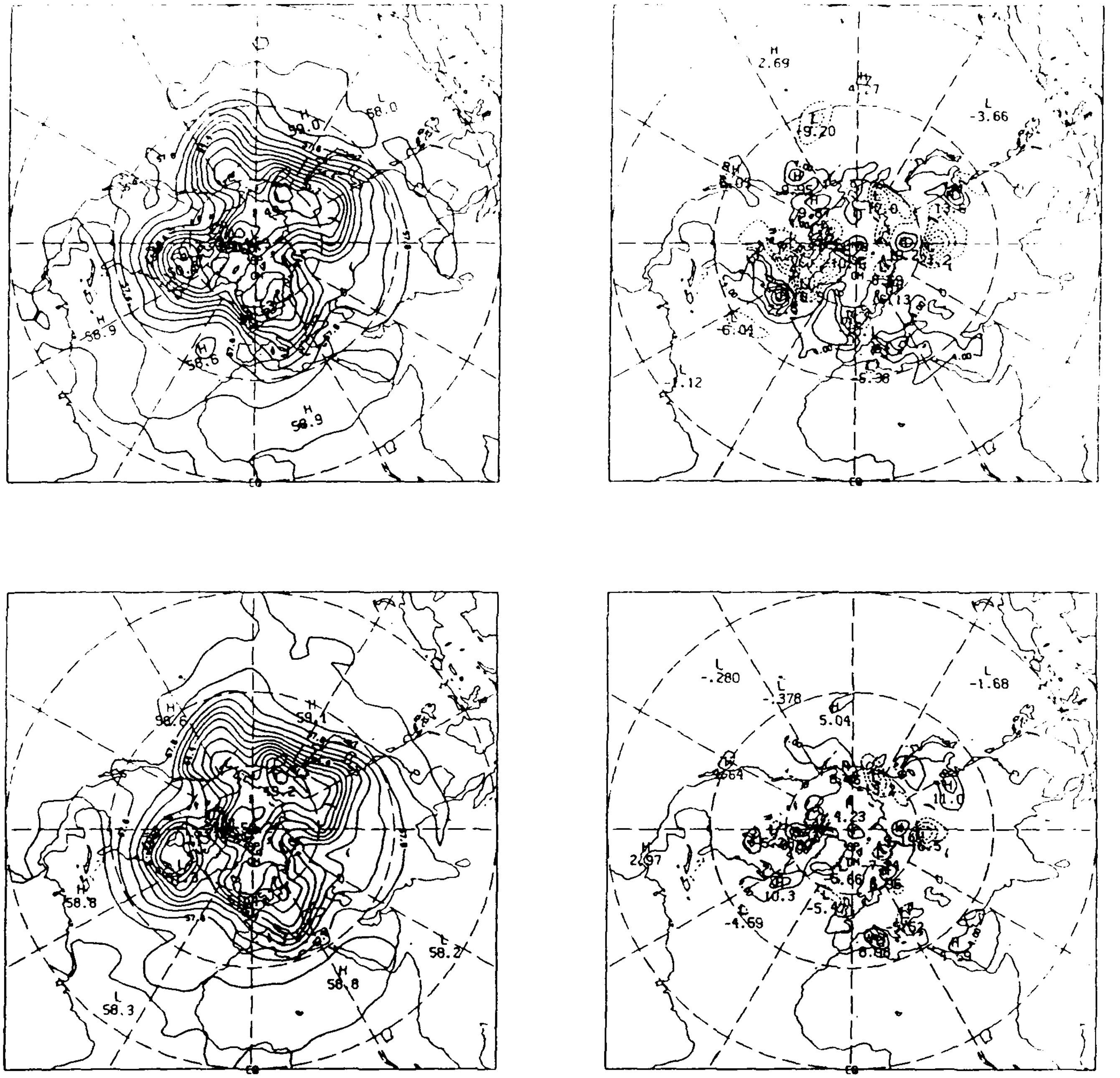
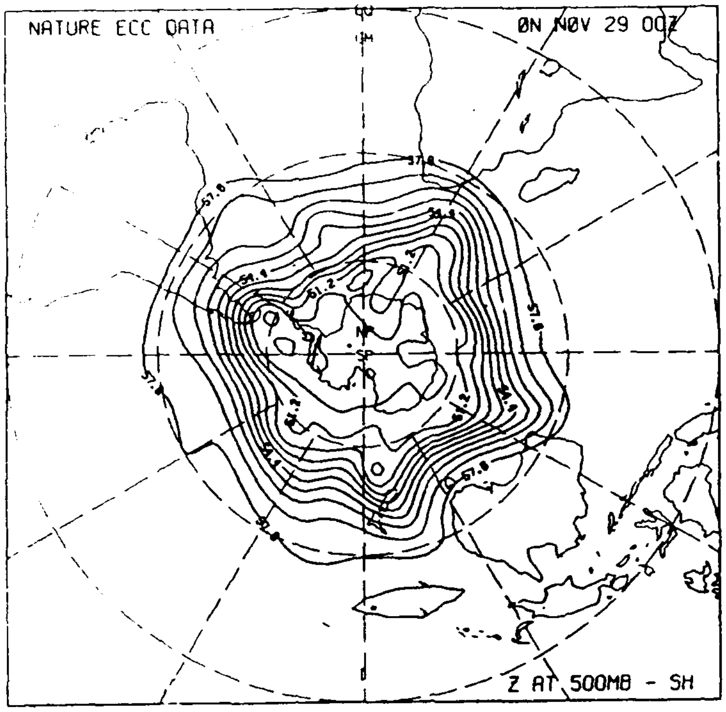


Fig. 5 The STATSAT (top) and WINDSAT (bottom) 500 mb Northern Hemisphere height field 48 hour forecasts (left) and 48 hour forecast errors (right) at 00 GMT 23 November 1979. Here and below the 500 mb height errors are displayed with a 40 m contour interval.



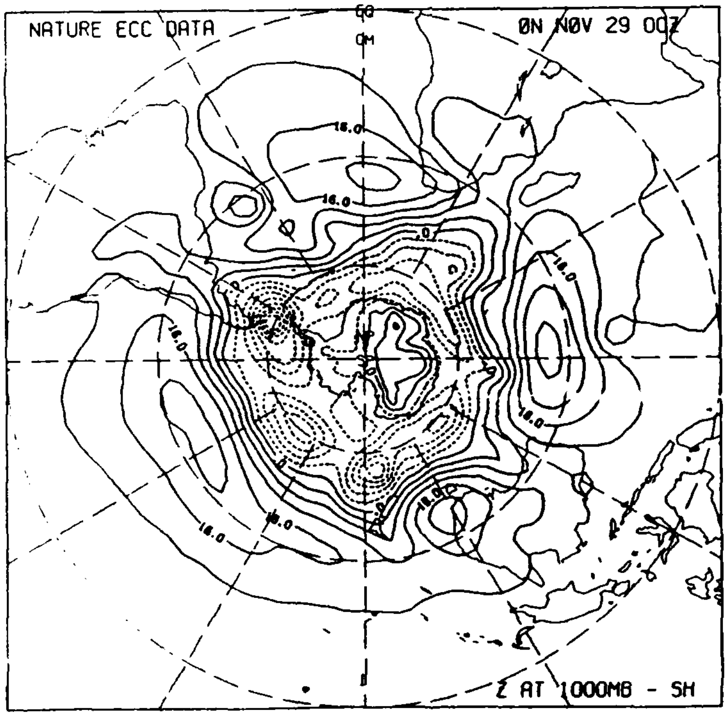


Fig. 6 The nature run 500 mb (top) and 1000 mb (bottom) Southern Hemisphere height field at 00 GMT 29 November 1979. Here and below the 1000 mb height field is displayed with a 40 m contour interval.

As another example of forecast skill, we consider the 96 hour height forecast in the Southern Hemisphere. Fig. 6 shows the 500 and 1000 mb height fields for the nature run. The corresponding 500 mb fields for STATSAT and WINDSAT are shown in Fig. 7. Again the WINDSAT errors are much smaller than the STATSAT errors. Note the trough SW of Australia is weaker and too far west in STATSAT, but is fairly well depicted by WINDSAT. Also the trough which is lined up with the southern part of South America has a slight phase error in WINDSAT but is split in STATSAT. The 1000 mb charts are more striking (Fig. 8). The WINDSAT errors in the Indian Ocean region are much smaller: Note the poor forecast of the large anticyclone in this region in the STATSAT analysis. WINDSAT does poorly in the Pacific sector. At the southern tip of South America, there is an intense cyclone with a minimum height of -320 m. This feature is forecast somewhat better by WINDSAT.

Although we have concentrated on the height field in the above description, the WINDSAT wind fields are much better that the STATSAT wind fields. As an example, Fig. 9 shows the 200 mb (T15 filtered) divergence fields from the nature run, STATSAT analysis and WINDSAT analysis at the end of the experiment. The patterns in the extratropics all agree quite well, however the amplitude is too high in WINDSAT and much too high in STATSAT. In the tropics the analyses are not very good.

4.2 Objective Statistical Evaluation

The conclusions of the previous section are borne out, in fact summarized by the objective statistics which we have calculated. In these comparisons we include some results of our Special Sensor Microwave (SSM) experiments for the purpose of comparison. These experiments are described in detail in a separate report (Grassotti et al., 1989). Here we examine some of the rms errors calculated as rms differences with respect to the nature run. Later, in Section 5 we describe a more detailed analysis of the forecast errors in terms of rms differences with respect to the radiosondes.

Fig. 10 shows the global 500 mb rms height error of the NOSAT, STATSAT, SSM and WINDSAT analyses and forecasts. The NOSAT analysis

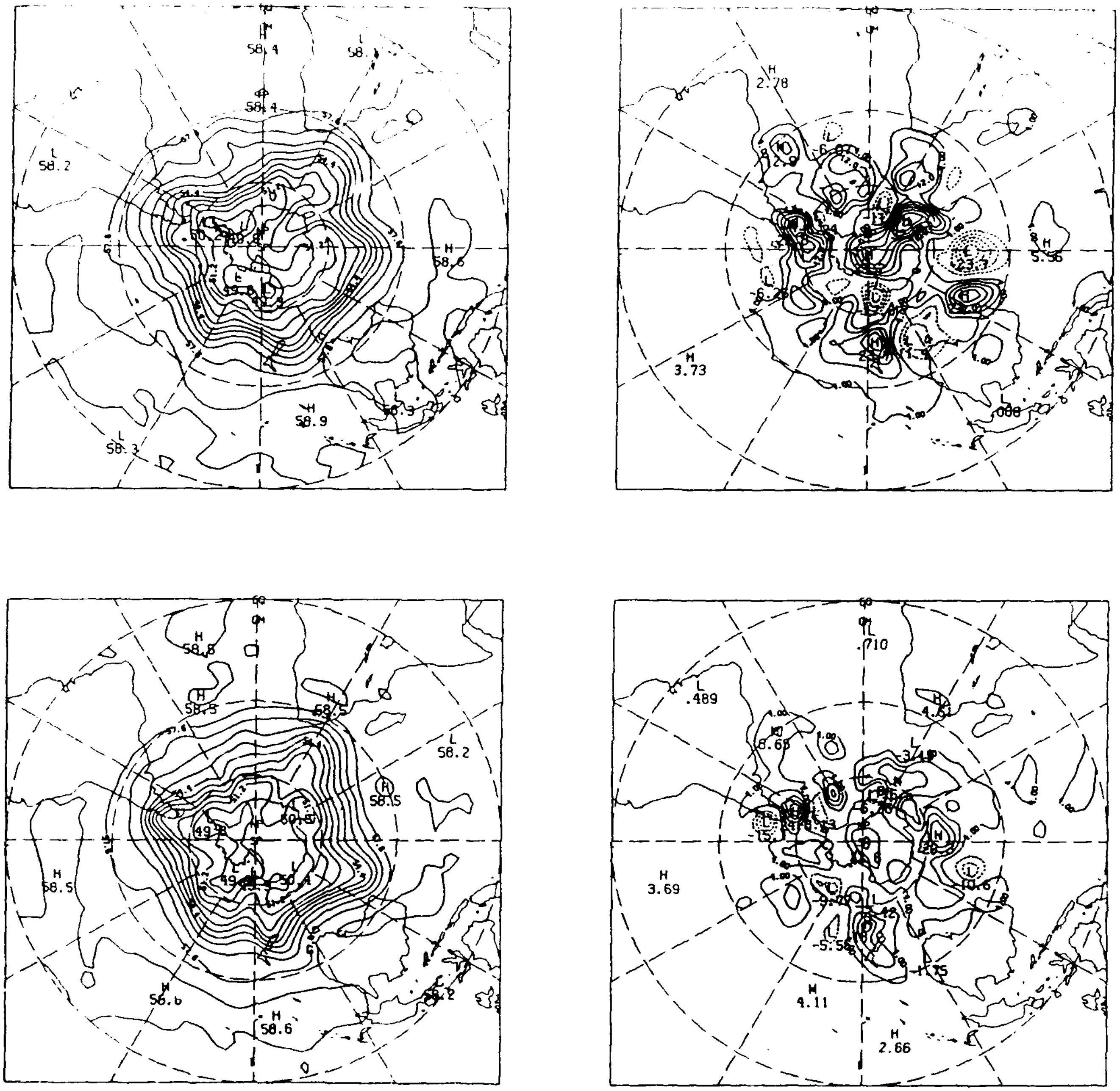


Fig. 7 The STATSAT (top) and WINDSAT (bottom) 500 mb Southern Hemisphere height field 96 hour forecasts (left) and 96 hour forecast errors (right) at 00 GMT 29 November 1979.

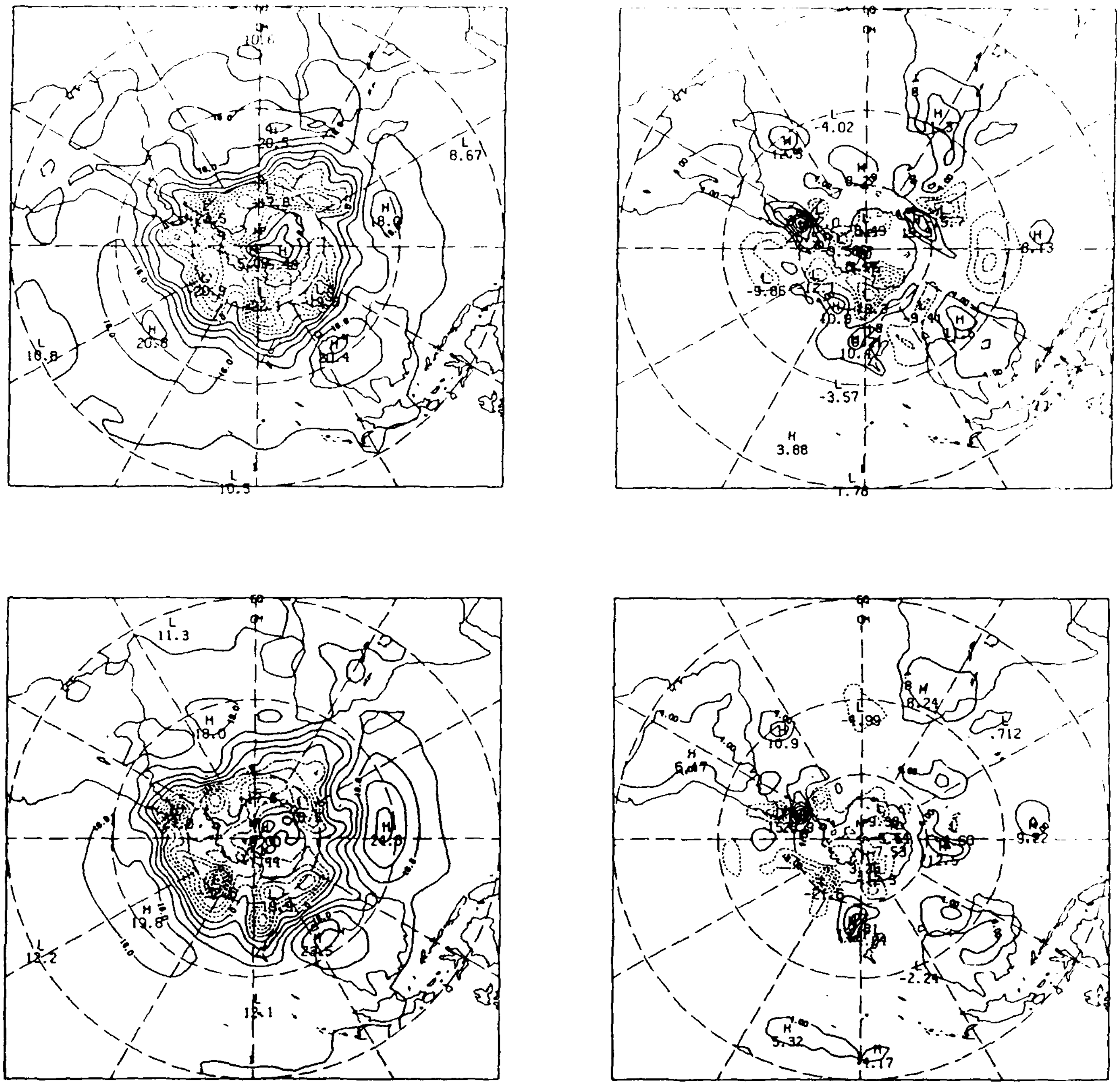


Fig. 8 The STATSAT (top) and WINDSAT (bottom) 1000 mb Southern Hemisphere height field 96 hour forecasts (left) and 96 hour forecast errors (right) at 00 GMT 29 November 1979. The 1000 mb height errors are displayed with a 40 m contour interval.

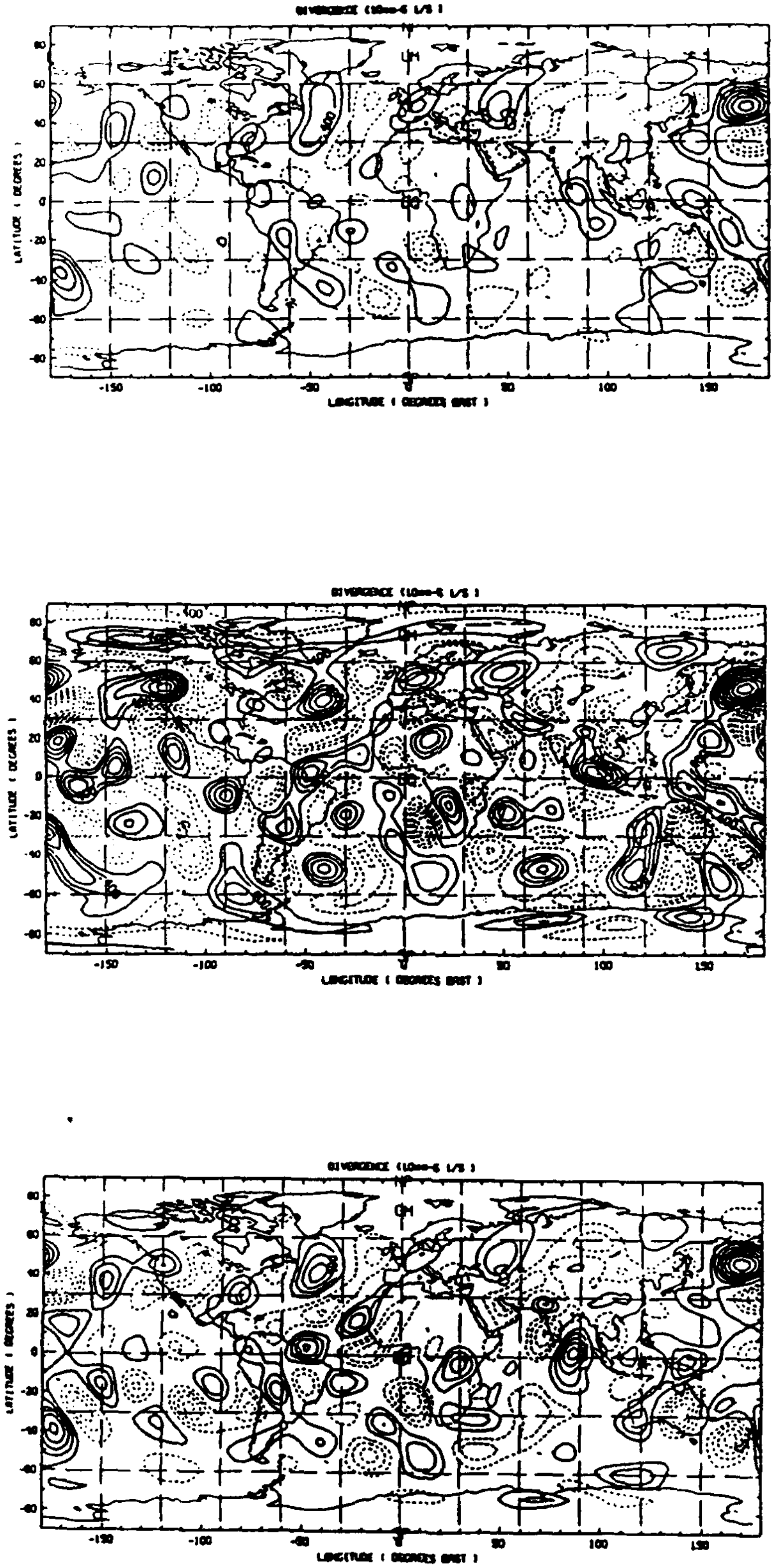


Fig. 9 The nature run (top), STATSAT (middle) and WINDSAT (bottom) 200 mb divergence fields for 00 GMT 25 November 1979. These fields have been spectrally truncated at T15. The contour interval is 2E-6 (1/s).

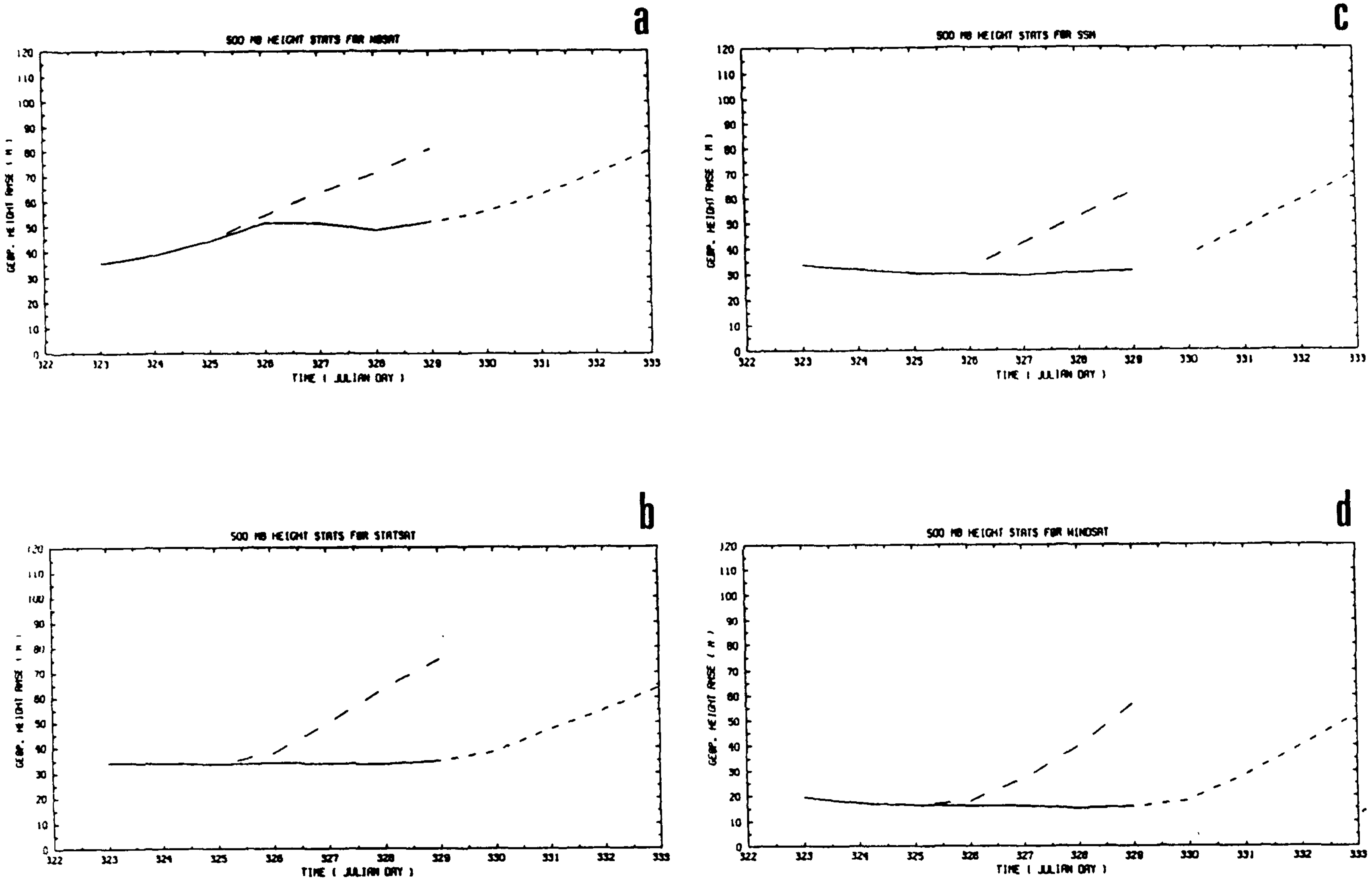


Fig. 10 Rms height errors at 500 mb. (a) NOSAT, (b) STATSAT, (c) SSM, (d) WINDSAT. Analysis errors are shown in solid curves, forecast errors in dashed curves. Julian day 322 corresponds to 00 GMT 18 November.

errors increase from the 35 m typical for the STATSAT analysis to 50 m by day 4 of the assimilation, whereas the STATSAT analysis errors decrease by 1-2 m over the assimilation period. This is an indication of how well the spinup process has performed. The forecast error growth is more rapid in STATSAT, but errors remain smaller than those of the NOSAT forecast for the length of the forecasts. The SSM analysis errors are consistently smaller than those of STATSAT, but by only 2-3 m. The WINDSAT data have a definite and dramatic impact on the analysis error; by 24 hours the error has dropped to 20 m and continues to slowly decline thereafter. The WINDSAT forecast errors, since they start from such good initial conditions are the smallest of all the experiments. The objective results thus confirm our impressions from the subjective evaluation of Section 4.1.

Results at other levels largely mirror those at 500 mb. The 1000 mb height statistics (not shown) show a much smaller impact of the satellite data (NOSAT analysis errors differ by no more than 3 m from STATSAT), but qualitatively the same results apply.

Considering the 850 mb relative humidity field (Fig. 11), we see that SSM provides the best analyses yet the best forecasts are obtained from WINDSAT. This is more so in the extratropics than the tropics; presumably the relative humidity forecasts are determined largely by the large scale fields of temperature and winds in the extratropics and the WINDSAT analyses of these are superior. SSM is always better than NOSAT which in turn is somewhat better than STATSAT. The particularly low growth rate of relative humidity errors for STATSAT is an indication that the errors have already saturated and that the analyses are nearly worthless.

4.3 Evaluation of Zonal Cross Sections

We examined zonal cross sections of u and v wind components, temperature and relative humidity at individual synoptic times and averaged over the last five 0000 GMT analyses of the experiments. Zonal averaging is denoted here by square brackets ([ ]) and time averaging by an overbar ($\bar{\ }$). Our purpose here is to determine how well the assimilation system is capturing the mean meridional circulation, pole to equator temperature and humidity structure and zonal jets and to describe the impact of the

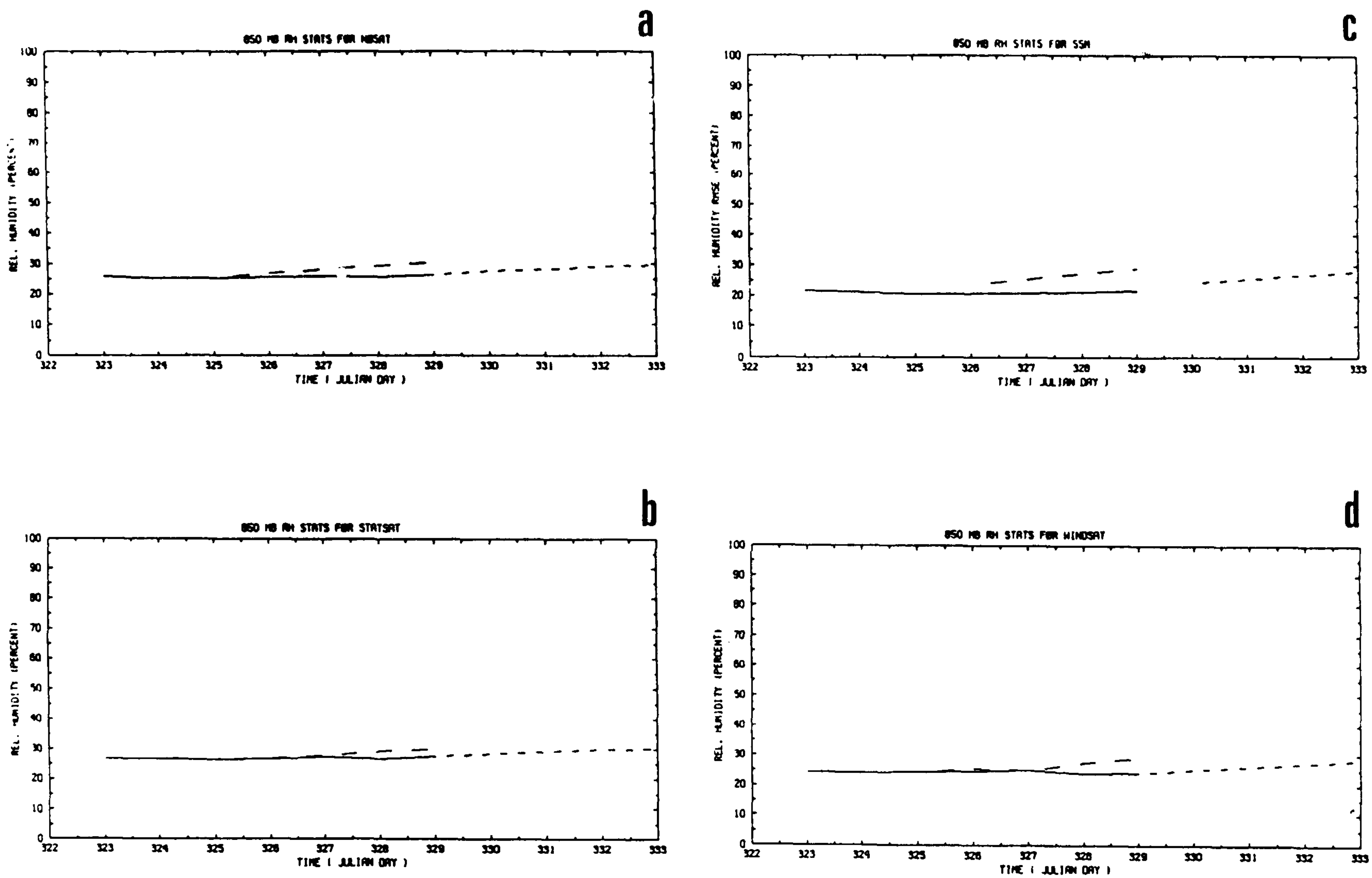


Fig. 11 Global rms analysis/forecast errors for 850 mb relative humidity. (a) NOSAT, (b) STATSAT, (c) SSM, (d) WINDSAT. Solid curves denote analysis, broken curves are forecasts.

different observing systems on these features. These features are important climate diagnostics. Errors in zonally averaged quantities are also important to NWP because errors in the climate make a persistent contribution to errors in the analyses and these errors may be useful in diagnosing faults in the physical parameterizations used in the model.

We concentrate here on the time averaged fields for the nature run and STATSAT and on the impact of WINDSAT on the wind, temperature and humidity fields. These fields and differences from the nature run are displayed in Figs. 12 through 15. For comparison the some results for the SSM experiment are also displayed.

Considering first the zonal time averaged zonal wind component (Fig. 12), we see in all cases the midlatitude jets peak near 200 mb. The Northern Hemisphere jet is somewhat narrower and stronger. The jet maximum is near 30 m/s in all cases. Easterly winds extend through the depth of the atmosphere in the tropics. The trade winds (surface easterlies) are a bit stronger in the Northern Hemisphere. The nature $[\bar{u}]$ is quite similar to observed fields. Compared to the GFDL monthly mean $[\bar{u}]$ for November 1979 (Lau, 1984), the jets in the nature run are positioned somewhat poleward and have slightly different magnitudes. STATSAT, SSM and WINDSAT $[\bar{u}]$ agree well with the nature run. In all three cases the errors are order 1 m/s with WINDSAT having somewhat smaller errors.

In Fig. 13, which depicts $[\bar{v}]$ both Hadley and Ferrel cells are readily apparent. On the other hand the surface southerlies in the Southern Hemisphere. Ferrel cell are stronger than in the Northern Hemisphere. The maximum convergence at the surface where the two returning branches of the Hadley cell meet is at 10° N. The corresponding upper level divergence is at 250 mb. The Northern Hemisphere Hadley cell appears to be somewhat stronger than the Southern Hemisphere Hadley cell. Compared to the GFDL monthly means, the surface winds agree fairly well, but the poleward branches of the Hadley cell are twice as strong in the GFDL analyses. Comparing STATSAT, SSM and WINDSAT to the nature run, we see that all three analyses are qualitatively in agreement with the nature run. Magnitudes of the $[\bar{v}]$ are all similar. However STATSAT has more small scale features in the tropopause region and SSM misses the strong

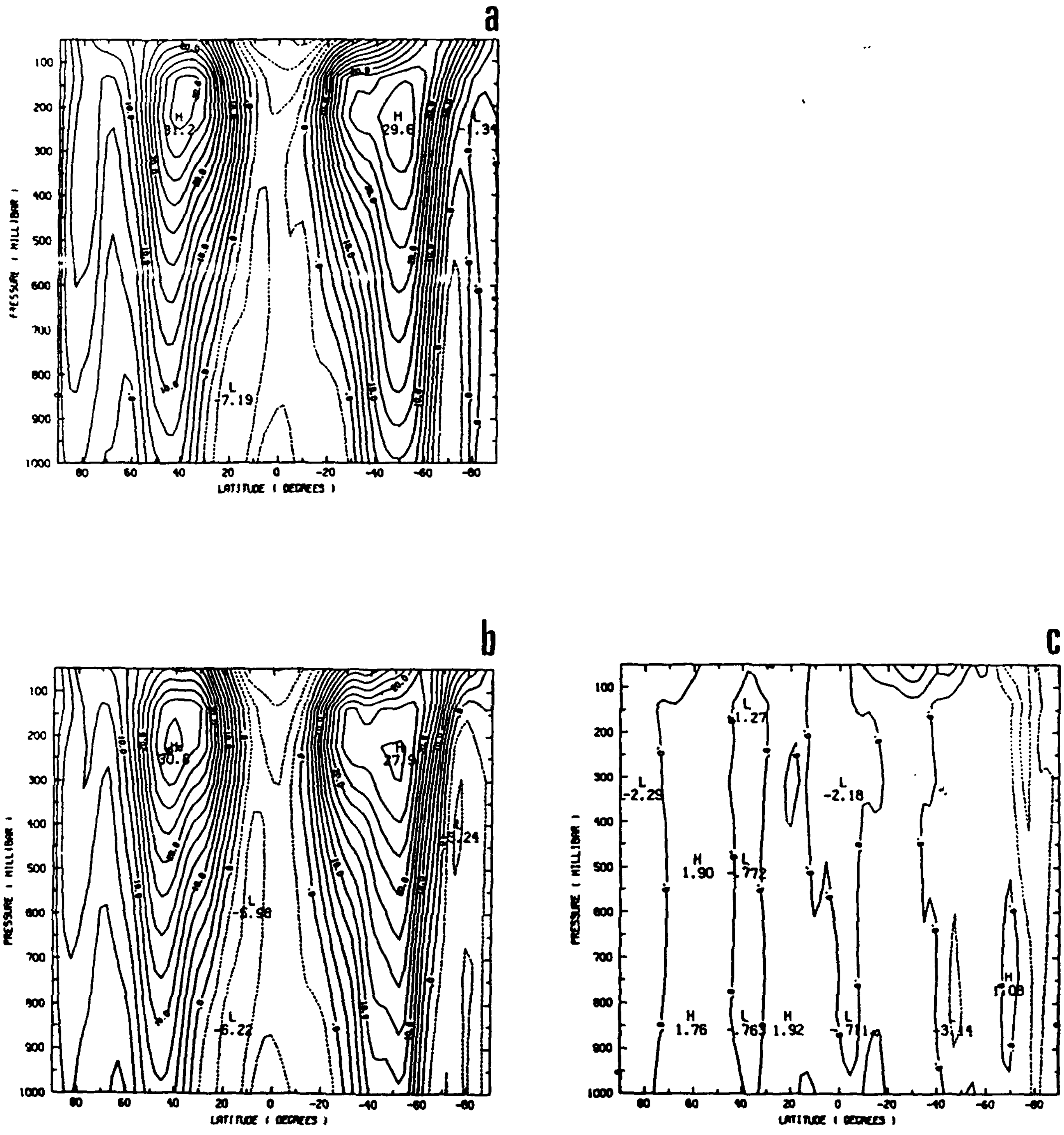


Fig. 12 Zonal time averaged u component of wind. (a) The nature run, (b) STATSAT, (c) STATSAT - the nature run, (d) WINDSAT, (e) WINDSAT - the nature run, (f) SSM, (g) SSM - the nature run. Contour interval is 2.5 ms^{-1}, negative values are dashed.

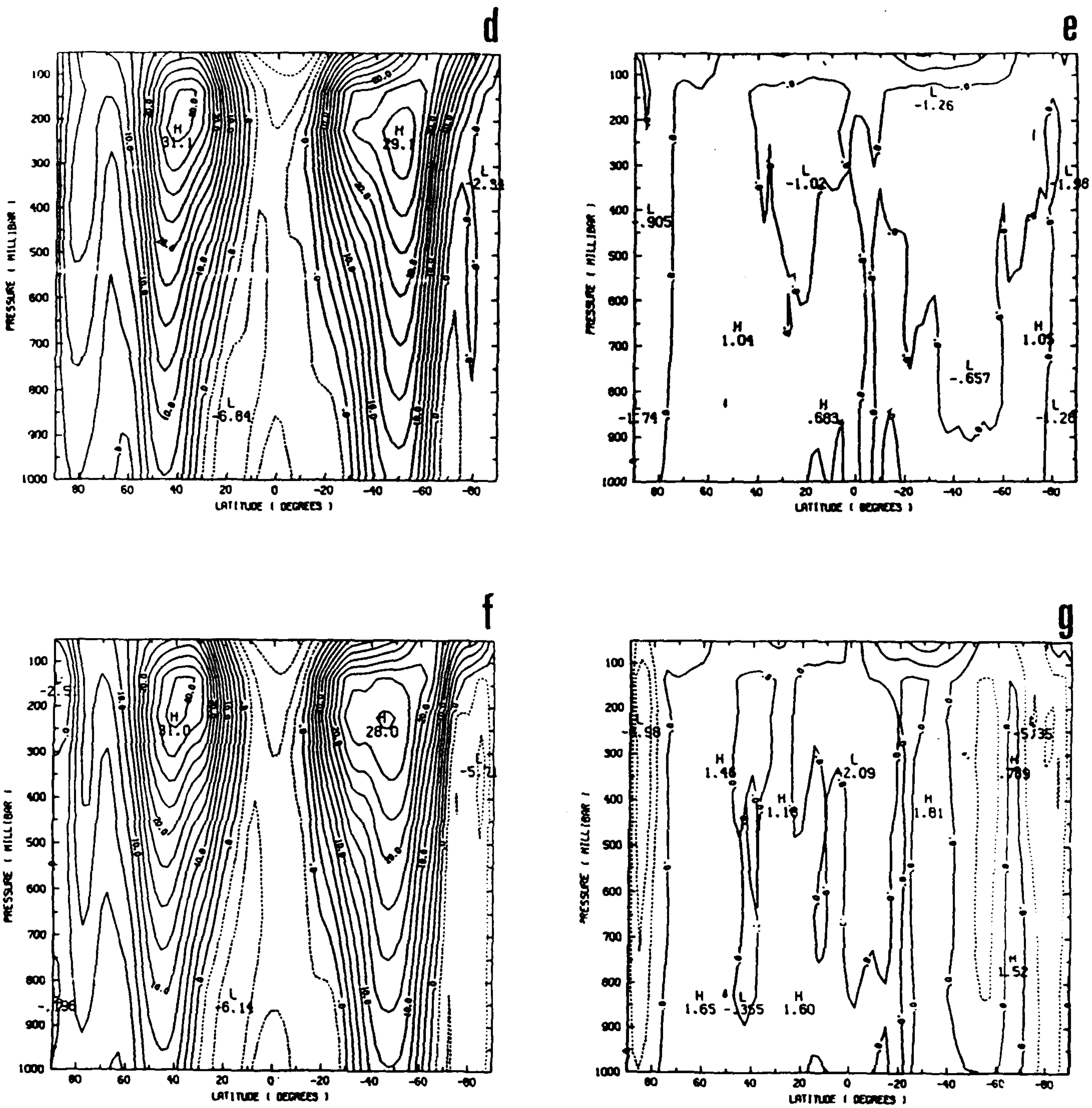


Fig. 12 (continued)

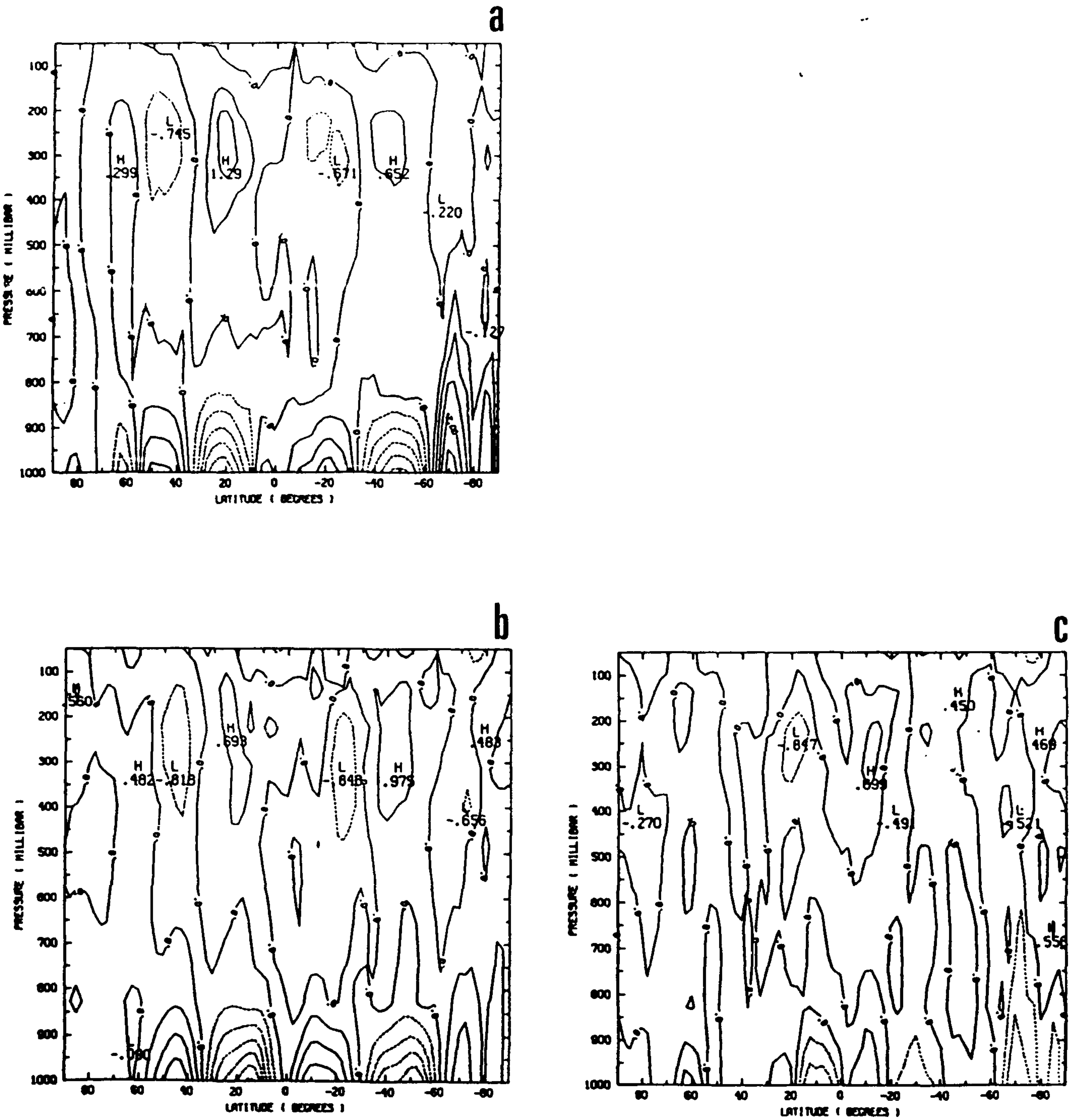


Fig. 13 Zonal time averaged v component of wind. (a) The nature run, (b) STATSAT, (c) STATSAT - the nature run, (d) WINDSAT, (e) WINDSAT - the nature run, (f) SSM, (g) SSM - the nature run. Contour interval is 0.5 ms^{-1}, negative values are dashed.

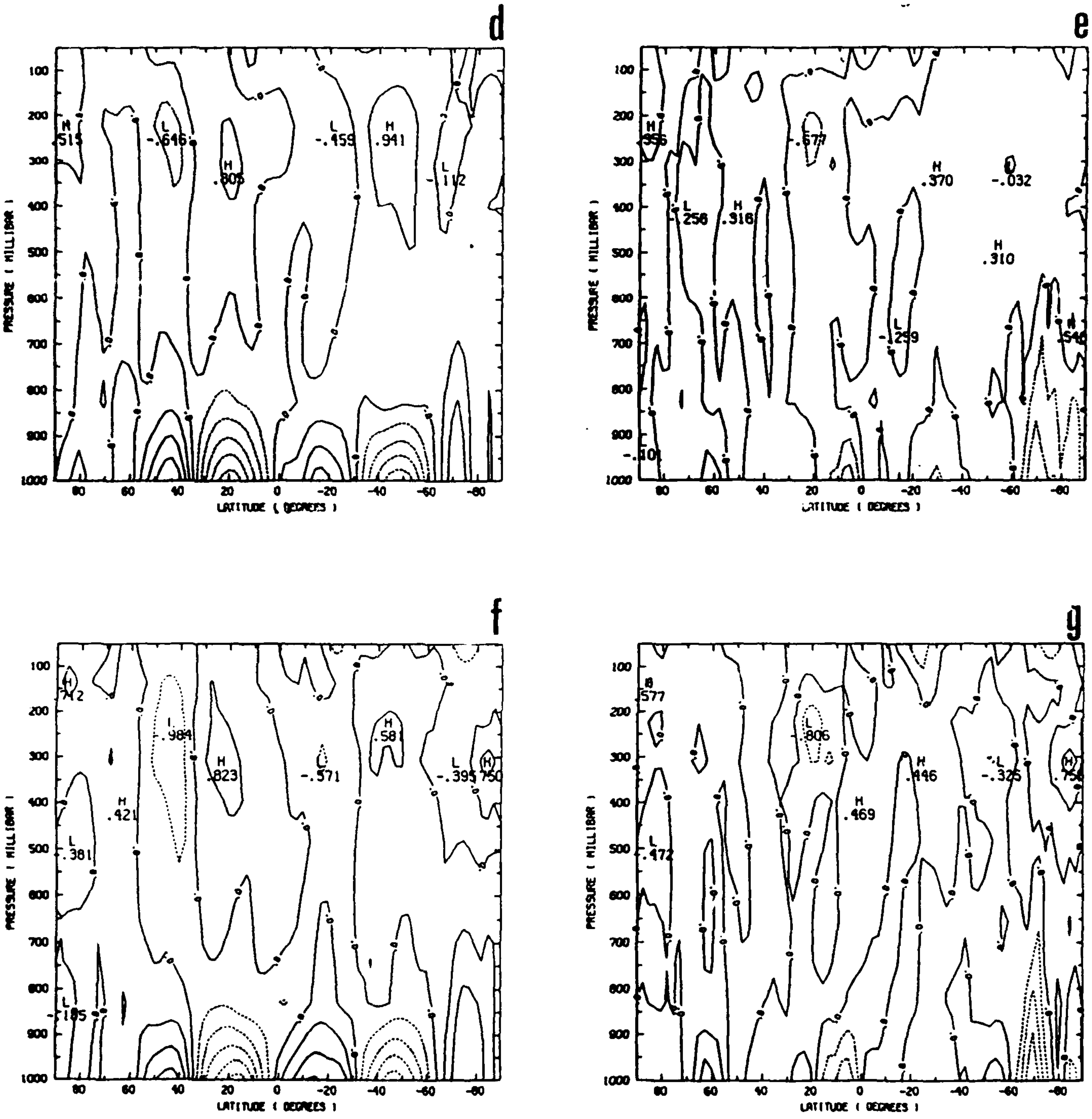


Fig. 13 (continued)

northerly surface flow over the Southern Ocean at 70° S. At the surface WINDSAT and STATSAT appear to be roughly equivalent.

Considering the amount of high quality wind data available to WINDSAT, the small improvements to the zonally averaged wind fields are disappointing. The lack of improvement in the tropical mean meridional circulation may be caused by deficiencies in the assimilation system. First, the analysis wind increments must be nondivergent: Although the height and wind analysis decouple in the tropics, the wind analysis still uses structure functions derived from the height structure function and the assumption of geostrophy. Secondly, in the NMI, there is no account of the effects of cumulus convection.

The $[\overline{T}]$ (Fig. 14) has a broad maximum between 20° S and 20° N at all levels through the tropopause. Poleward of 20°, temperature decreases. The magnitude of the poleward temperature gradient decreases with elevation up to the 200 mb level where the gradient reverses. Tropopause height varies from 200 mb in the polar regions to 50 mb at the equator. The South Pole tropopause is very cold (only 200° K). The nature run $[\overline{T}]$ agrees well with the GFDL values except that the GFDL tropical tropopause is roughly at the 100 mb level. Compared to the nature run, we see that STATSAT is too cold (by 2 K) in the tropical planetary boundary layer (below about 900 mb) and too warm at the equator at 850 mb (by 1.4 K). This implies the equatorial region is too stable. In the Southern Hemisphere between 60° S and 80° S STATSAT is too cold below 850 mb (up to -5 K) and there are large postive errors over Antarctica. Note that topography is at about 700 mb south of 80° S. In the Artic below 850 mb STATSAT is also too warm. There are large errors above the tropopause; the poles are warm and the equatorial region is cold. WINDSAT has errors similar to STATSAT, however the upper level errors are substantially reduced in magnitude, the equatorial planetary boundary layer is even colder (by 3 K) and the errors over Antartica are reduced.

The $[\overline{RH}]$ (Fig. 15) in the nature run is very moist at the surface where it has a maximum of 85% at 10° N and a minimum of 75% at 30° N. Moisture is carried upwards by the ascending Hadley circulation near the equator and dry air is brought downward by the descending branches of the

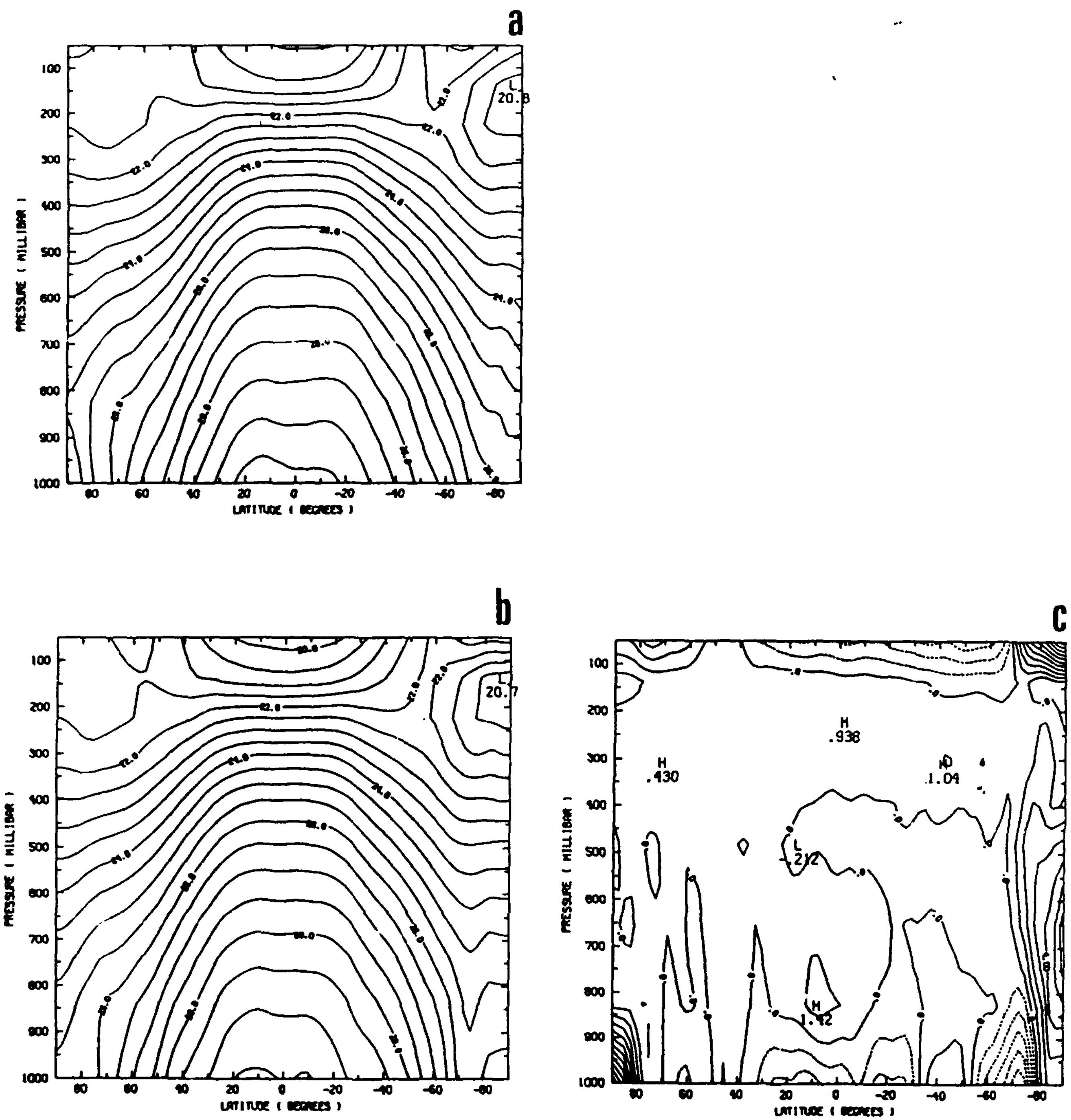


Fig. 14 Zonal time averaged temperature. (a) The nature run, (b) STATSAT, (c) STATSAT - the nature run, (d) WINDSAT, (e) WINDSAT - the nature run, (f) SSM, (g) SSM - the nature run. In analyses, values (in K) have been multiplied by 0.1, contour interval 0.5 (K/10). In differences values are unscaled, contour interval is 1K.

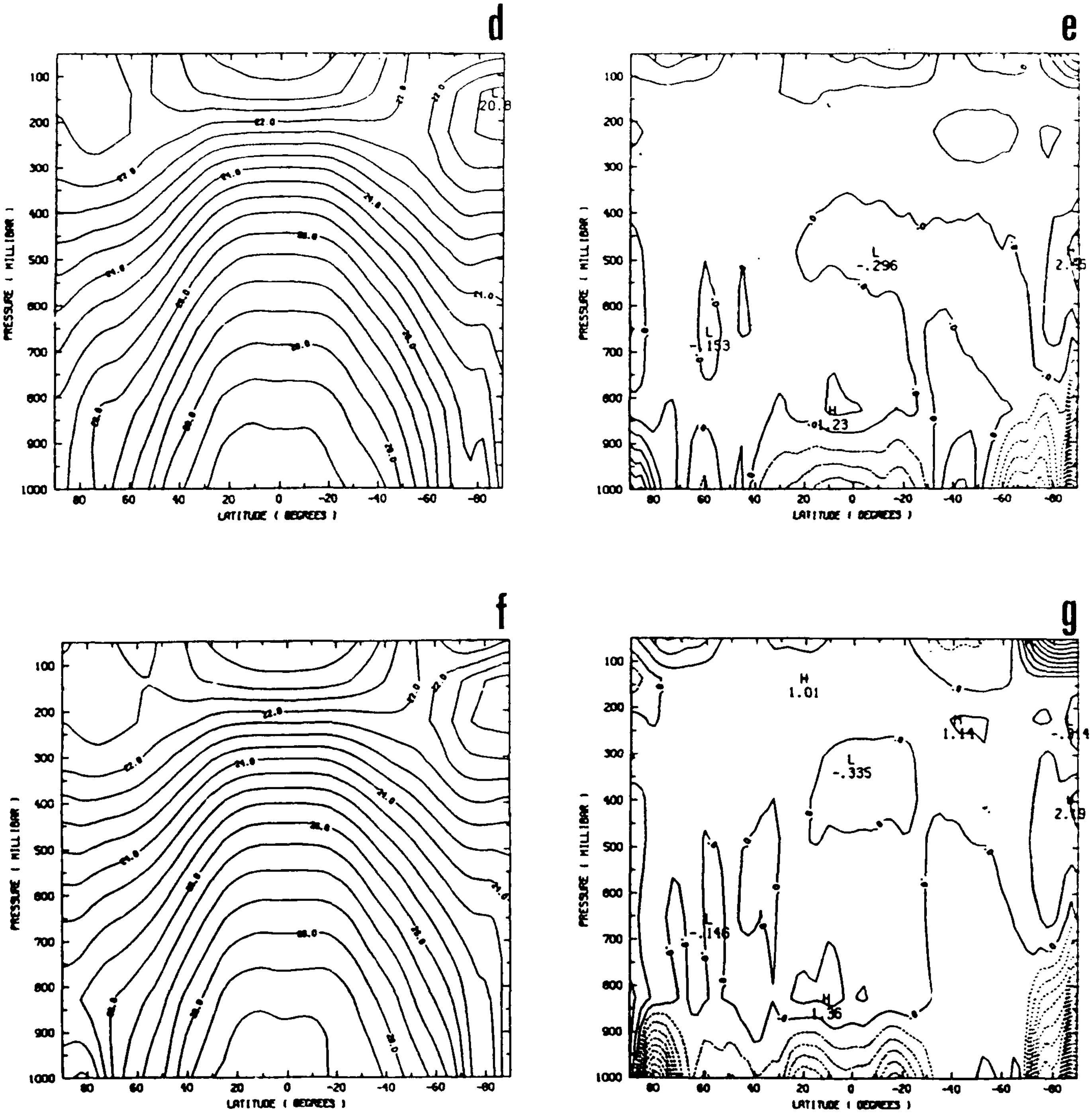


Fig. 14 (continued)

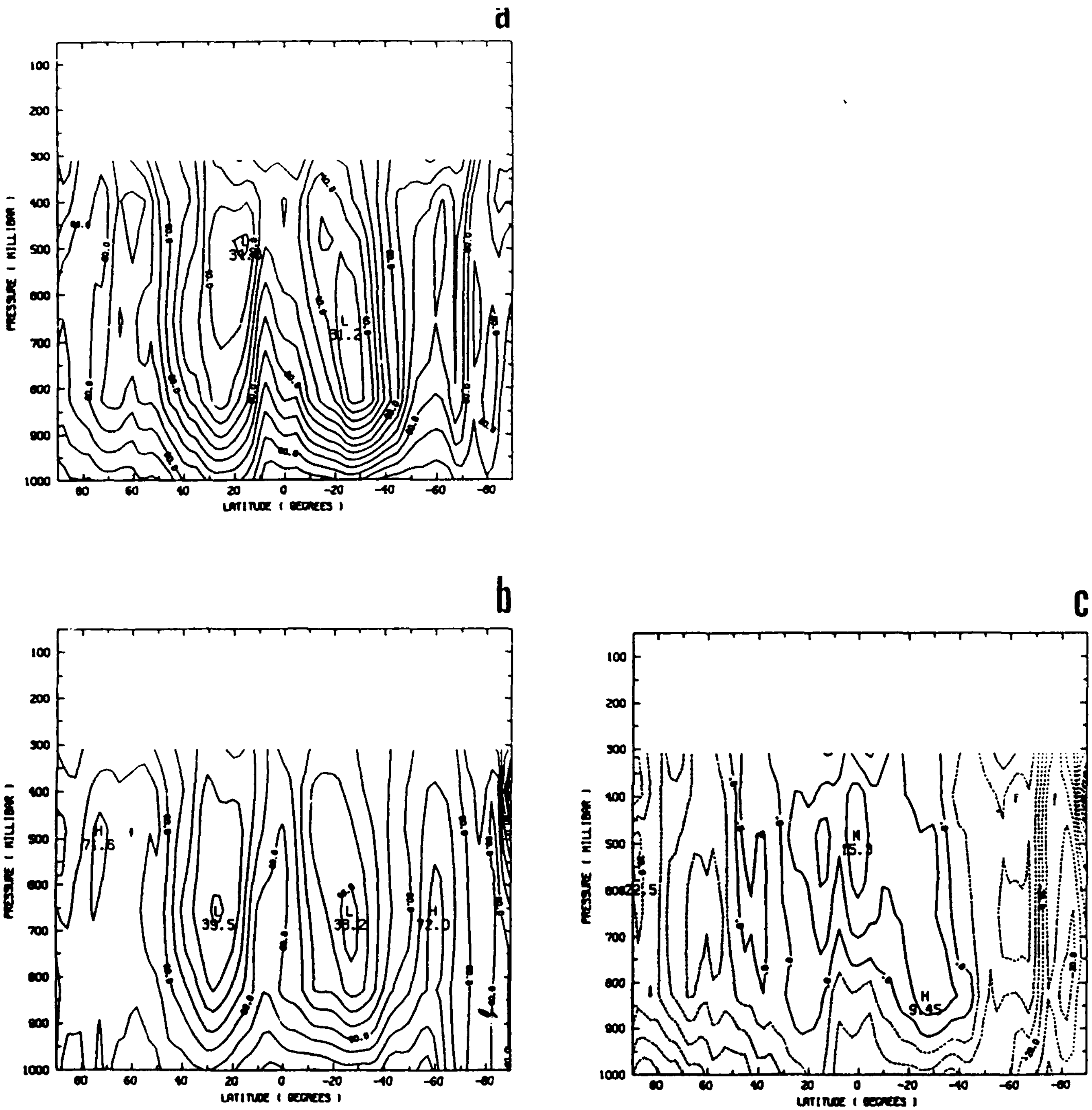


Fig. 15 Zonal time averaged relative humidity. (a) The nature run, (b) STATSAT, (c) STATSAT - the nature run, (d) WINDSAT, (e) WINDSAT - the nature run, (f) SSM, (g) SSM - the nature run. Contour interval is 5 percent, negative values are dashed.

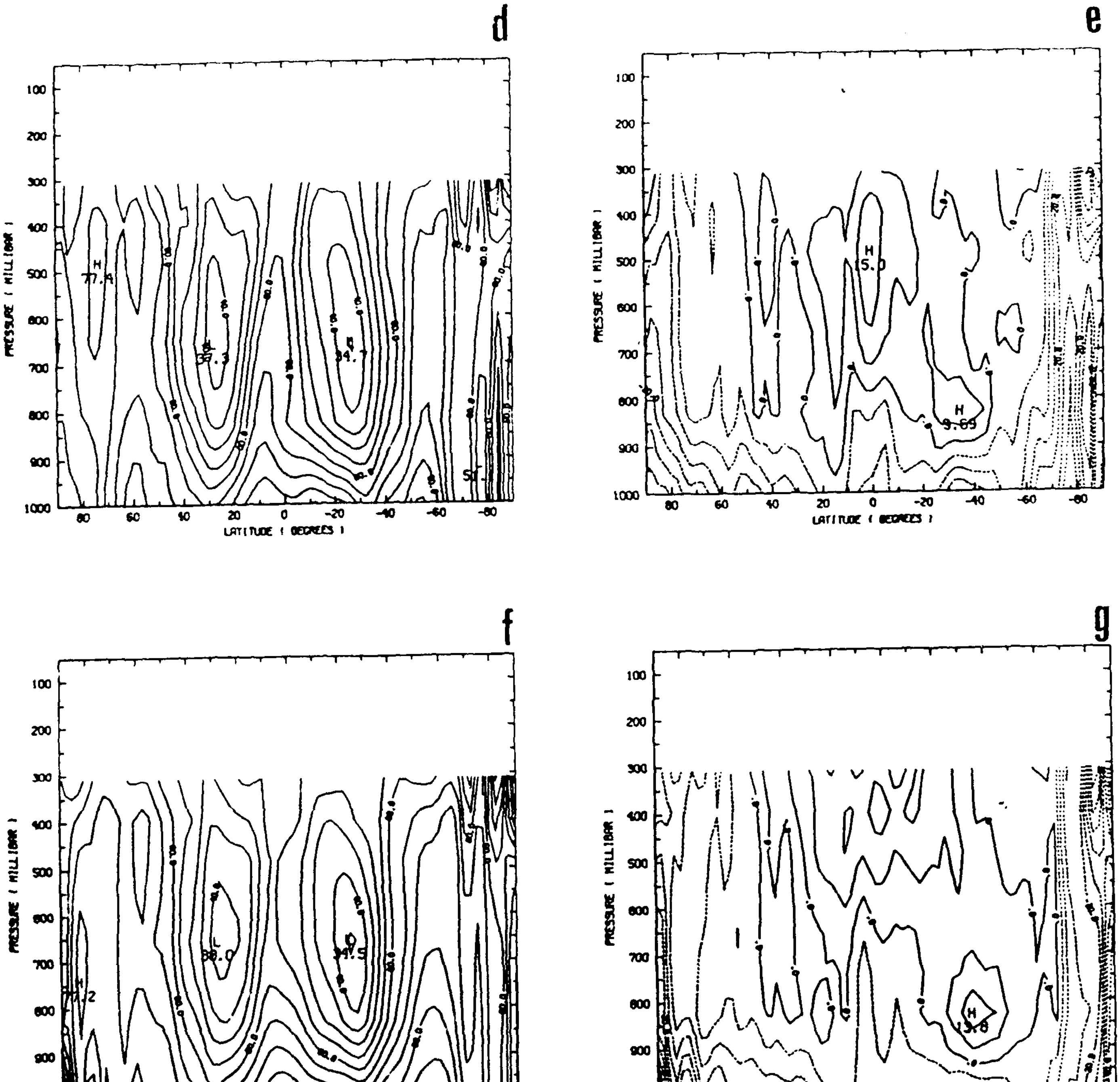


Fig. 15 (continued)

Hadley cell near 30° N and S. This pattern is repeated, but with smaller amplitude by the Ferrel cells. $[\overline{RH}]$ tends to decrease from the surface to 500 mb and then increase again up to 300 mb. At 500 mb the minimum $[\overline{RH}]$ is about 30%. $[\overline{RH}]$ errors are quite large in all assimilations. Typically the polar regions and boundary layer are too dry and the mid latitude and tropical atmosphere above the planetary boundary layer is too moist. Since moist air is more buoyant these errors tend to stabilize the analyses. Both the temperature and humidity errors in the tropics suggest that too much stabilizing convection may be taking place in the AFGL model.

5. OSSE Forecast Errors and Their Calibration

A luxury of OSSEs is the ability to exactly compute error measures. We took advantage of this in our discussion of the analyses. Analysis errors in the real world are not well known. In fact, recently Daley and Mayer (1986) presented analysis error of the OSSE experiments of Atlas et al. (1985) which were discussed in the introduction as surrogates for real analysis errors. We now turn to an examination of the forecast errors. In order to have a closer correspondence with the real world and to simplify our calibration procedure we have calculated rms differences between the forecasts and the simulated radiosondes. We then describe a procedure to calibrate these differences using the NOSAT - STATSAT impact observed in our previous OSEs as a yardstick, and present some results of the calibration procedure. Some discussion of the results of the companion OSSE experiments, SSM and SSM+TOVS (WINDSAT), are included here for comparison.

5.1 Rms Differences between OSSE Forecasts and Radiosondes

Rms differences between OSSE forecasts and the simulated radiosondes used in the data assimilation experiments were calculated for different regions and for several variables at each layer in the atmosphere. The variables examined include geopotential height, temperature, vector wind, relative humidity, and cloud cover. Mean differences were also calculated and examined but were small relative to the rms differences for nearly all variables. Some regularities observed in the mean differences are described in Section 5.4. One aspect of the procedure we used is that the

forecast heights are anchored by the verifying radiosonde report. Consequently, height errors described here are actually thickness errors. As described below we curve fit the data to determine impacts in terms of predictability time, i.e. the length of the useful forecast.

Due to the variable density of radiosonde coverage, global averages are very similar to Northern Hemisphere extratropical averages and Southern Hemisphere extratropical averages are based on fairly small samples. Of course, these statistics are biased towards land areas. Consequently, differences between NOSAT and STATSAT are less pronounced than they might otherwise be. As a result of these factors as well as the radiosonde errors themselves, we expect only qualitative agreement between impacts described here and impacts measured in Section 4.2 by comparing forecasts grid point by grid point to the nature run.

Examples of the growth of forecast errors as evidenced by the rms differences are displayed in Figs. 16 through 18. In each figure there are three panels, showing results averaged over three forecasts for the Northern Hemisphere extratropics, tropics and Southern Hemisphere extratropics. Here the tropics are taken to run from 30 S to 30 N. Rms difference curves for smaller regions which were studied generally behaved as described here for the larger parent regions. In the Southern Hemisphere, the rms difference curves sometimes exhibit a sawtooth pattern due to sampling problems; these are usually about 60 RAOBs at 00 GMT and only about 40 at 12 GMT in the Southern Hemisphere. Most of the non-reporting RAOBs are in the Australian sector.

Considering first the rms differences for 500 mb geopotential (Fig. 16), we see that impacts in the Northern Hemisphere are relatively small. WINDSAT leads STATSAT by approximately 7 hours. Midway between these two lie SSM and NOSAT. SSM+TOVS is equivalent to STATSAT. In the tropics, the differences are all near 20 m and the growth rate is very small. Apparently, the tropical analysis errors are so large that the error growth is near saturation. Note that the standard deviation of 500 mb geopotential in the tropics is typically 20 m (Oort, 1983). At least during the initial part of the forecast WINDSAT has a slight edge, which because of the slow growth of differences corresponds to 9 or 10 hours of

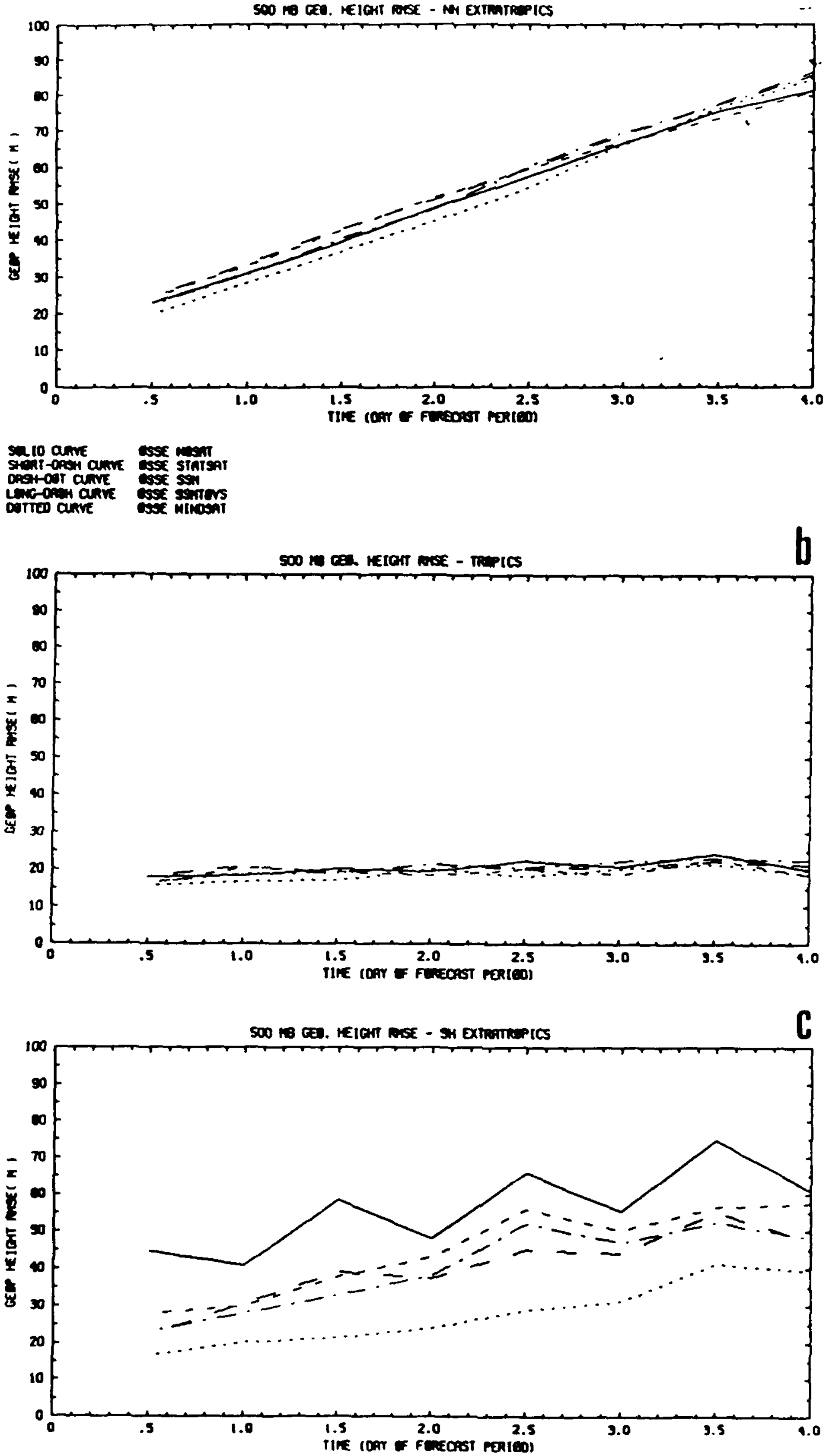


Fig. 16 Forecast rms error growth, 500 mb height. (a) Northern Hemisphere extratropics, (b) Tropics, (c) Southern Hemisphere extratropics.

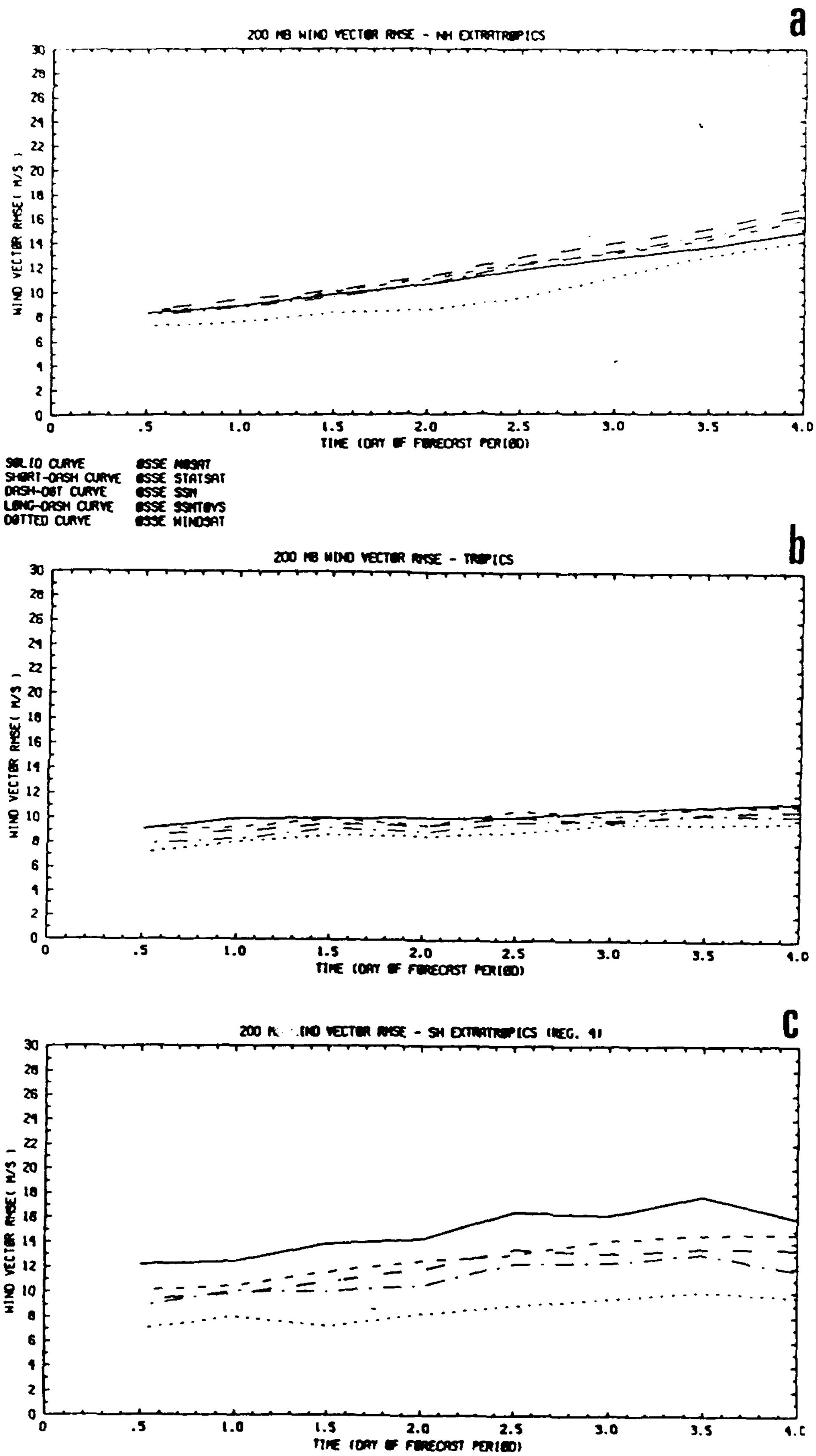


Fig. 17 Forecast rms error growth, 200 mb wind vector. (a) Northern Hemisphere extratropics, (c) Southern Hemisphere extratropics.

predictability time. Impacts in the Southern Hemisphere are very large. WINDSAT is 36 hours better than STATSAT, which is in turn more than 36 hours better than the NOSAT forecasts.

Next, we consider the rms vector wind differences at 200 mb (Fig. 17). Clearly, WINDSAT always yields a big improvement. Compared to STATSAT, WINDSAT provides 1, 2 and 2.75 day improvements in forecast skill in the Northern Hemisphere tropics and Southern Hemisphere respectively. NOSAT is particularly poor over the Southern Hemisphere and the tropics. Of the three sounder systems, SSM is generally better, improving predictability by at least a day in the Southern Hemisphere and tropics.

The rms differences for relative humidity at 850 mb are shown in Fig. 18. Impacts in terms of forecast time are all relatively small in the Northern Hemisphere generally in the range 3 to 12 hours. In the extratropics, WINDSAT is the best. This might have been anticipated since the WINDSAT analyses of mass and wind are superior in the extratropics and since relative humidity is so strongly influenced by the large scale synoptic systems which are better forecast by WINDSAT. In the tropics, the SSM moisture forecasts are the best. Overall the ranking is WINDSAT, SSM, NOSAT, SSM+TOVS and STATSAT. It appears that using TOVS degrades the moisture analysis. Since TOVS relative humidity retrievals were not used, this poses a conundrum. TOVS data affect the model specific humidity indirectly because the analyzed variables are temperature and relative humidity. That is, in an area with only TOVS data the updated temperature field is combined with the unaltered relative humidity field to update the model specific humidity.

We also calculated rms differences in cloud cover layer by layer. Cloud cover was diagnosed from relative humidity using the inverse Tibaldi scheme (as described in Section 3.5). Invariably, the corresponding relative humidity and cloud cover plots look very similar. Fig. 19 shows the global rms differences for 850 mb relative humidity and cloud cover. Except for the fact that the cloud cover errors are larger in magnitude, the curves in the two panels of Fig. 19 are nearly the same. For this reason we have not included any other cloud cover rms difference statistics in this report.

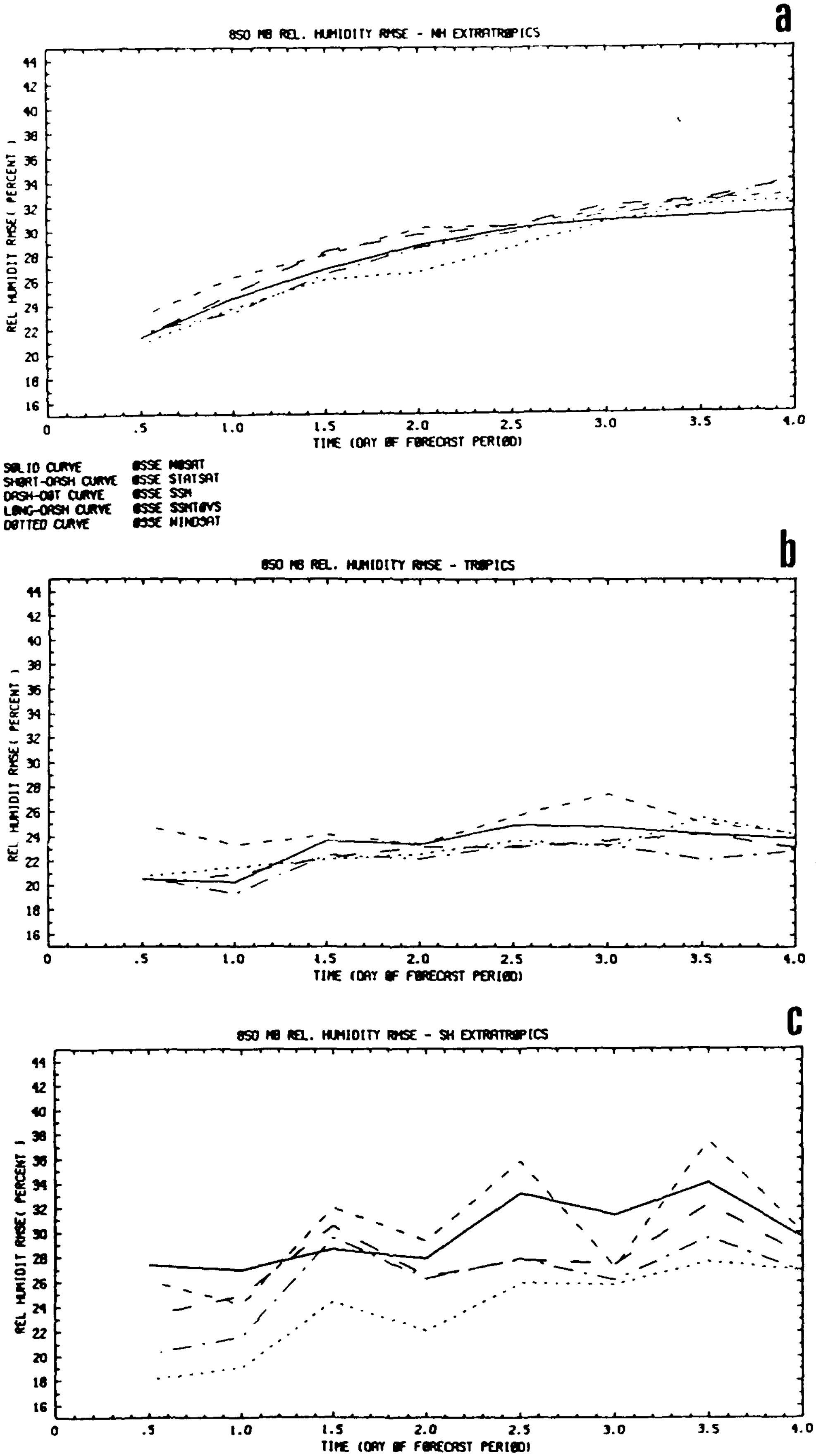


Fig. 18 Forecast rms error growth, 850 mb relative humidity. (a) Northern Hemisphere extratropics, (b) Tropics, (c) Southern Hemisphere extratropics.

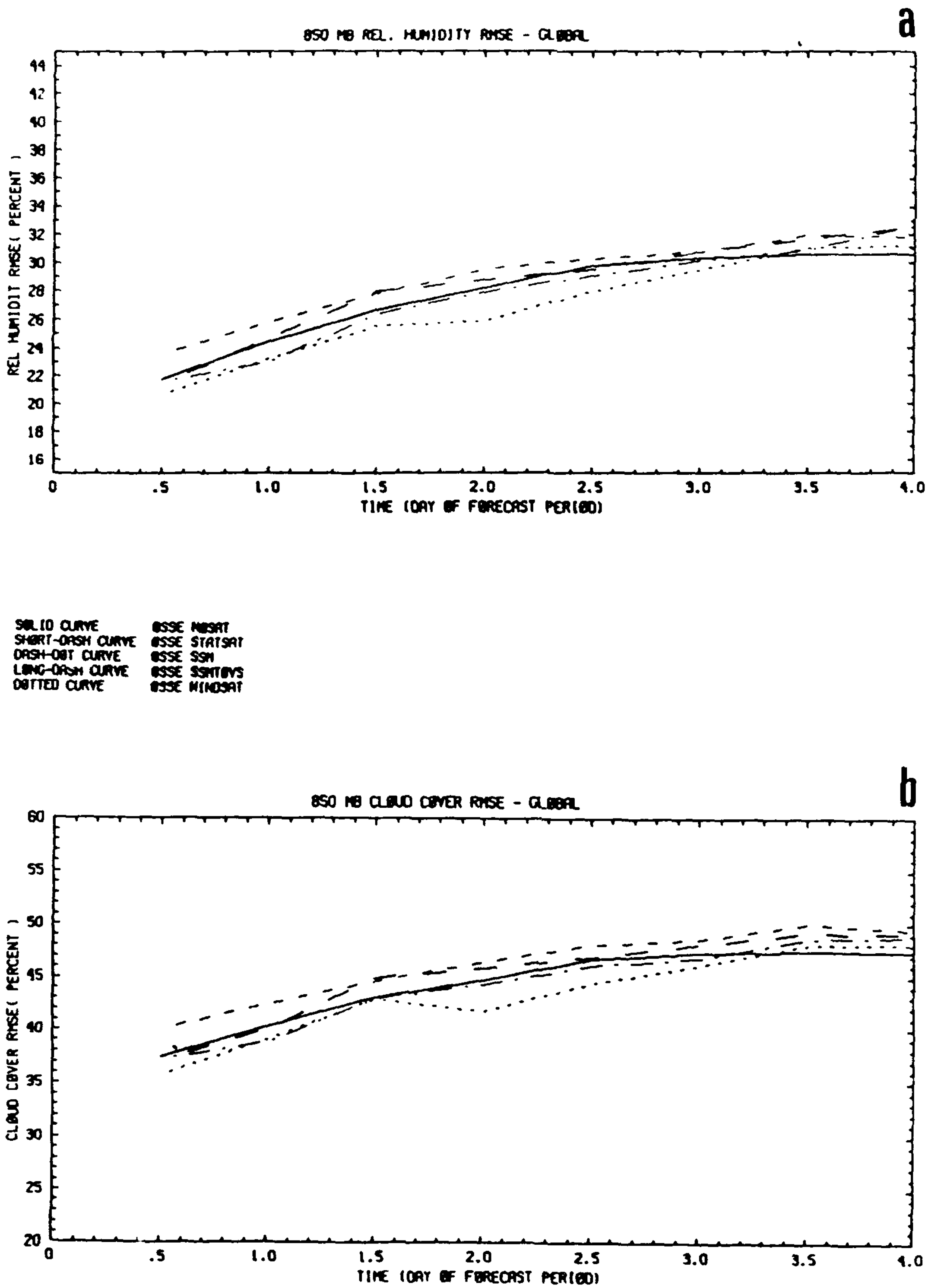


Fig. 19 Forecast rms error growth. (a) 850 mb relative humidity, global statistics, (b) 850 mb cloud cover, global statistics.

We examined the growth of rms differences for levels others than those described here. The results shown here are generally representative. In the next section tables summarizing all the levels are given.

5.2 Calibration procedure

It has been observed that OSSE forecasts are too good because any two models, such as the model used in the experiments and the model used to generate nature, are more alike than any model and the real atmosphere. Consequently, it is unwise to naively carry over the forecast impacts observed in OSSEs to the real world. For example, at short forecast times, OSSE forecasts tend to be so good that there is little room for improvement; adding a new observing system might then have a smaller impact than in the real world. On the other hand, at longer forecast times, real data forecasts will be so bad that a new observing system will have no impact while the corresponding OSSE impact may be significant.

For these reasons it is desirable to calibrate the OSSE results. However, for the present experiments we find that the OSSE impacts are fairly similar to the OSE impacts and the calibration procedure does not greatly alter the conclusions one might draw from the OSSE results directly. To minimize practical and interpretive difficulties we use only very simple approaches. Our principal assumption is that the OSE impact of adding or removing an observing system is proportional to the corresponding OSSE impact. In our calibrations we always take STATSAT to be our standard. We use the NOSAT - STATSAT difference to determine the constant of proportionality.

Impact may be measured in many ways. Useful impact measures should account for differences in variability from season to season and from region to region. For example an impact of 10 geopotential meters is meaningless without the context of place, season and vertical level. Usually, for the purpose of comparison, it is reasonable to scale the squared errors by their respective climate variance.

It is also often helpful to define impact in terms of the change in predictability time. For example, one might define the predictability time as the time at which the mean squared forecast error reaches the

climate variance level. Forecasts with errors this large are normally worthless. A positive impact in predictibility time would then indicate the additional time that the forecast remains useful.

Measuring impact in terms of predictability time is especially useful when combined with a commmon idealization of the growth of forecast error. A number of simple parameterizations of the growth of error have been advanced (e.g. Dalcher and Kalnay, 1987 and references therein). Remarkably good fits to ensemble averaged forecast error growth curves have been obtained, by fitting relatively simple autonomous constant coefficient ordinary differential equations. These coefficients describe the growth of small errors, the saturation of large errors at the climate variance level and the source of errors due to modeling deficiences. Since these constants should be the same for a set of experiments, e.g. for all our February OSEs, all fitted error growth curves for the experiments should be the same except for a translation with respect to the time axis. This shift is the impact in terms of predictability time.

A direct reading of the predictability times from the rms difference curves proved difficult because our sample is rather small. We could fit the parameterization of Dalcher and Kalnay (1987). Instead, we took advantage of the observation that our rms difference curves grow nearly linearly, at least during the forecast period from 12 to 48 hours, to fit these data with a series of straight lines having a common slope. In the Northern Hemisphere these fits were very good. They are less reliable in the Southern Hemisphere and tropics, presumably because the number of radiosondes in these regions is small. Our results for height, wind and relative humidity are shown in Tables 3 through 5. In the tables, each item in the columns labeled R**2 is the fraction of variance explained by the fitting procedure or equivalently the square of the correlation coefficient. The other columns in the tables display the predictability time impacts of the various experiments, all relative to STATSAT. Some of these values have been quoted earlier. These impacts are the difference in the x (or time) intercepts of the fitted lines. (These intercepts are proportional to the y (or rms differences) intercepts with proportionality factor equal to the common slope.)

Table 3. Predictability impacts (days) for height (a), wind (b) and relative humidity (c) for the N.H. extratropics. Impacts for cases with R**2 less than 0.25 (i.e. for correlations less than 0.5) are not shown.

A. Height

l	P(l)	November OSSE R**2	November OSSE NOSAT	November OSSE SSM/T	November OSSE SSM+TOVS	November OSSE WINDSAT	February OSE R**2	February OSE NOSAT	June OSE R**2	June OSE NOSAT
1	1000	.000	-	-	-	-	.603	.266	.219	-
2	850	.822	.183	-.143	.022	.601	.975	-.287	.920	-.319
3	700	.991	.168	.118	.011	.488	.996	-.048	.989	-.032
4	500	.999	.141	.144	-.008	.307	.999	.023	.985	-.003
5	400	.999	.098	.113	-.017	.255	.998	.022	.960	-.020
6	300	.998	.038	.064	-.010	.186	.998	.000	.978	-.070
7	250	.996	.019	.050	-.013	.167	.998	-.019	.958	-.115
8	200	.996	-.011	.045	.012	.103	.998	-.051	.874	-.145
9	150	.994	-.013	.060	.029	.120	.997	-.104	.778	-.175
10	100	.996	-.100	.049	.065	-.112	.993	-.121	.449	-.444
11	70	.999	-.412	.060	.214	-.294	.990	-.127	.111	-
12	50	.998	-.699	.064	.307	-.368	.972	-.220	.434	-.585

B. Vector wind

l	P(l)	November OSSE R**2	November OSSE NOSAT	November OSSE SSM/T	November OSSE SSM+TOVS	November OSSE WINDSAT	February OSE R**2	February OSE NOSAT	June OSE R**2	June OSE NOSAT
1	1000	.969	.414	.191	-.094	.928	.909	.161	.910	.451
2	850	.963	.474	.207	-.210	.915	.987	.122	.977	-.144
3	700	.981	.458	.149	-.314	.973	.990	.040	.932	-.184
4	500	.976	.410	.179	-.176	.866	.996	.015	.948	-.160
5	400	.985	.373	.182	-.145	.846	.995	.009	.984	-.152
6	300	.980	.233	.134	-.154	.806	.977	-.058	.990	-.103
7	250	.970	.214	.147	-.163	.941	.989	-.060	.987	-.135
8	200	.948	.126	.214	-.142	1.079	.979	-.171	.990	-.235
9	150	.860	.001	.318	-.095	1.301	.965	-.292	.988	-.250
10	100	.892	.232	.489	.127	1.244	.929	-.175	.984	-.108
11	70	.884	.265	.572	.331	1.222	.893	-.035	.964	.010
12	50	.929	.245	.506	.399	1.001	.938	-.076	.970	.157

C. Relative humidity

l	P(l)	November OSSE R**2	November OSSE NOSAT	November OSSE SSM/T	November OSSE SSM+TOVS	November OSSE WINDSAT	February OSE R**2	February OSE NOSAT	June OSE R**2	June OSE NOSAT
1	1000	.781	.213	.197	-.013	.490	.822	-.154	.107	-
2	850	.959	.320	.388	.168	.543	.935	.100	.964	.233
3	700	.954	.365	.330	-.029	.719	.856	.136	.945	.120
4	500	.906	.475	.392	.242	1.121	.973	.400	.934	.262
5	400	.921	.161	.116	.014	.763	.990	.114	.886	-.019
6	300	.853	.188	.146	.066	.908	.861	.355	.846	-.086

Table 4. Predictability impacts (days) for height (a), wind (b) and relative humidity (c) for the tropics. Impacts for cases with R**2 less than 0.25 (i.e. for correlations less than 0.25) are not shown.

A. Height

l	P(l)	November OSSE R**2	November OSSE NOSAT	November OSSE SSM/T	November OSSE SSM+TOVS	November OSSE WINDSAT	February OSE R**2	February OSE NOSAT	June OSE R**2	June OSE NOSAT
1	1000	.000	-	-	-	-	.158	-	.312	.181
2	850	.131	-	-	-	-	.126	-	.643	-.466
3	700	.652	-.490	-.033	.176	.445	.484	.395	.779	-.641
4	500	.551	-.224	-.145	-.623	.610	.028	-	.241	-
5	400	.569	-.348	-.302	-1.007	.398	.695	.066	.672	-.804
6	300	.381	-.687	-.364	-.917	.414	.633	.658	.545	-.800
7	250	.447	-.499	.116	-.050	.465	.000	-	.880	-.921
8	200	.411	-.147	.731	.665	.940	.001	-	.881	-.517
9	150	.191	-	-	-	-	.041	-	.962	-.348
10	100	.171	-	-	-	-	.130	-	.556	-.354
11	70	.009	-	-	-	-	.094	-	.005	-
12	50	.005	-	-	-	-	.002	-	.000	-

B. Vector wind

l	P(l)	November OSSE R**2	November OSSE NOSAT	November OSSE SSM/T	November OSSE SSM+TOVS	November OSSE WINDSAT	February OSE R**2	February OSE NOSAT	June OSE R**2	June OSE NOSAT
1	1000	.450	.802	.815	.458	1.389	.047	-	.641	.323
2	850	.846	.416	.082	.029	.846	.813	-.246	.791	-.304
3	700	.881	.771	.765	.479	1.307	.368	.968	.883	.165
4	500	.550	.882	.956	.267	1.682	.402	.023	.770	-.321
5	400	.565	.861	1.150	.664	2.000	.334	.472	.775	-.794
6	300	.494	.261	1.762	-.128	3.908	.806	-.433	.537	-.575
7	250	.398	.418	1.454	.693	3.479	.788	-.212	.638	-.089
8	200	.664	-.462	1.317	.482	2.057	.666	-.649	.345	-.449
9	150	.739	.168	.932	.350	2.624	.328	1.527	.653	-.433
10	100	.033	-	-	-	-	.577	-.536	.902	-.158
11	70	.873	.355	.854	.636	1.719	.015	-	.736	-.267
12	50	.698	1.160	2.006	1.534	3.665	.049	-	.018	-

C. Relative humidity

l	P(l)	November OSSE R**2	November OSSE NOSAT	November OSSE SSM/T	November OSSE SSM+TOVS	November OSSE WINDSAT	February OSE R**2	February OSE NOSAT	June OSE R**2	June OSE NOSAT
1	1000	.114	-	-	-	-	.256	-1.478	.522	.319
2	850	.424	1.520	2.104	1.775	1.671	.521	-.121	.944	.071
3	700	.433	.601	.349	.460	1.018	.830	.210	.816	-.285
4	500	.795	-.237	.064	.086	.149	.000	-	.899	-.098
5	400	.624	-.135	-.192	-.243	-.059	.545	1.828	.393	.016
6	300	.157	-	-	-	-	.467	-.682	.480	-.275

Table 5. Predictability impacts (days) for height (a), wind (b) and relative humidity (c) for the S.H. extratropics. Impacts for cases with R**2 less than 0.25 (i.e. for correlations less than 0.5) are not shown.

A. Height

l	P(l)	November OSSE					February OSE		June OSE	
		R**2	NOSAT	SSM/T	SSM+TOVS	WINDSAT	R**2	NOSAT	R**2	NOSAT
1	1000	.000	-	-	-	-	.062	-	.024	-
2	850	.486	-.082	2.098	2.082	1.954	.000	-	.308	-.363
3	700	.795	-1.076	.469	.471	1.357	.582	-1.396	.318	-1.958
4	500	.671	-1.597	.498	.285	1.686	.678	-1.231	.688	-1.372
5	400	.668	-1.634	.524	.323	1.491	.469	-1.625	.777	-1.783
6	300	.734	-1.709	.472	.350	1.106	.399	-1.592	.751	-2.208
7	250	.742	-1.659	.528	.448	.949	.570	-1.423	.711	-3.210
8	200	.747	-1.605	.619	.502	.784	.403	-1.765	.598	-4.398
9	150	.717	-1.768	.626	.518	.798	.369	-1.745	.689	-4.280
10	100	.647	-2.243	.318	.267	.193	.483	-1.111	.560	-5.290
11	70	.577	-3.387	-.004	.841	-1.169	.802	-.787	.013	-
12	50	.610	-4.762	-.845	1.256	-2.376	.764	-.337	.184	-

B. Vector wind

l	P(l)	November OSSE					February OSE		June OSE	
		R**2	NOSAT	SSM/T	SSM+TOVS	WINDSAT	R**2	NOSAT	R**2	NOSAT
1	1000	.145	-	-	-	-	.057	-	.304	-1.757
2	850	.716	-.518	1.674	1.119	3.372	.795	-1.274	.688	-2.197
3	700	.696	-1.133	1.117	.406	2.559	.265	-2.042	.700	-2.487
4	500	.651	-1.435	.644	.172	2.306	.672	-.651	.320	-2.112
5	400	.601	-1.615	.536	.445	2.571	.573	-.714	.454	-1.461
6	300	.852	-1.547	.333	.081	2.055	.242	-	.154	-
7	250	.819	-1.359	.135	.043	1.516	.244	-	.024	-
8	200	.808	-1.562	1.011	.539	2.749	.777	-1.113	.004	-
9	150	.520	-1.716	1.448	.390	3.460	.117	-	.016	-
10	100	.051	-	-	-	-	.016	-	.124	-
11	70	.265	-1.454	1.621	.503	4.040	.059	-	.575	-3.541
12	50	.169	-	-	-	-	.290	-.335	.418	-1.083

C. Relative humidity

l	P(l)	November OSSE					February OSE		June OSE	
		R**2	NOSAT	SSM/T	SSM+TOVS	WINDSAT	R**2	NOSAT	R**2	NOSAT
1	1000	.133	-	-	-	-	.005	-	.705	-.075
2	850	.432	.062	1.102	.510	2.193	.364	-.969	.140	-
3	700	.741	.398	.832	.314	1.263	.188	-	.067	-
4	500	.322	-.026	2.434	.407	2.980	.045	-	.662	.225
5	400	.427	.014	1.667	.293	2.778	.493	-.037	.005	-
6	300	.168	-	-	-	-	.158	-	.705	.664

The NOSAT impact, i.e., the difference between NOSAT and STATSAT in the OSE experiments were then used to calibrate the OSSE results according to

$$(\text{Expected OSE impact}) = \frac{(\text{NOSAT OSE impact})*(\text{OSSE impact})}{(\text{NOSAT OSSE impact})}$$

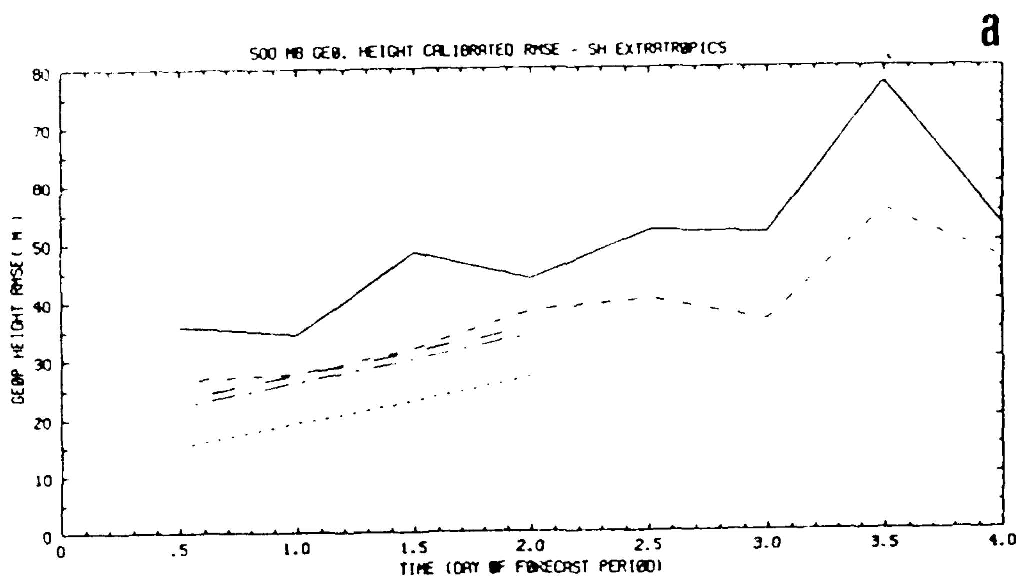
This provided us with calibrated intercepts which we combined with the observed OSE error growth rate (the common slope) to create calibrated rms difference curves for WINDSAT, SSM and SSM+TOVS experiments. These are displayed along with the observed OSE results in the figures. The horizontal distances between the various curves are the predictability time impacts.

5.3 Calibration Results

The calibration procedure described above allows us to translate our OSSE results into anticipated real world impacts in a quantitative manner. Several examples are provided in Figs. 20 through 22. Complete details are provided by Tables 3 through 5 and the above equation defining the expected impact.

In Fig. 20, the Southern Hemisphere 500 mb height rms differences for the STATSAT and NOSAT OSE experiments are plotted. The calibrated OSSE results for SSM, SSM+TOVS and WINDSAT are also plotted. These are the three straight lines between 12 and 48 hours on the plots. They are plotted only for this period since it is only this period which was used in the curve fitting. Note that the February (a) and June (b) OSEs provide two independent calibrations. In both cases the dramatic improvement seen in the OSSEs for WINDSAT is expected to carry over in actuality. A 36 hour improvement in forecast skill relative to STATSAT is anticipated and the expected impact of the SSM data is 12 hours. In the Northern Hemisphere (not shown) STATSAT and NOSAT OSSE results are nearly equivalent, so impacts expected from advanced observing systems cannot be calibrated.

The tropical 200 mb vector wind rms differences are shown for February (a) and June (b) OSSEs (Fig. 21). Because the growth rates for the



SOLID CURVE — NOSAT OSE
SHORT-DASH CURVE — STATSAT OSE
DASH-DOT CURVE — SSM CALIBRATION
LONG-DASH CURVE — SSMTOVS CALIBRATION
DOTTED CURVE — WINDSAT CALIBRATION

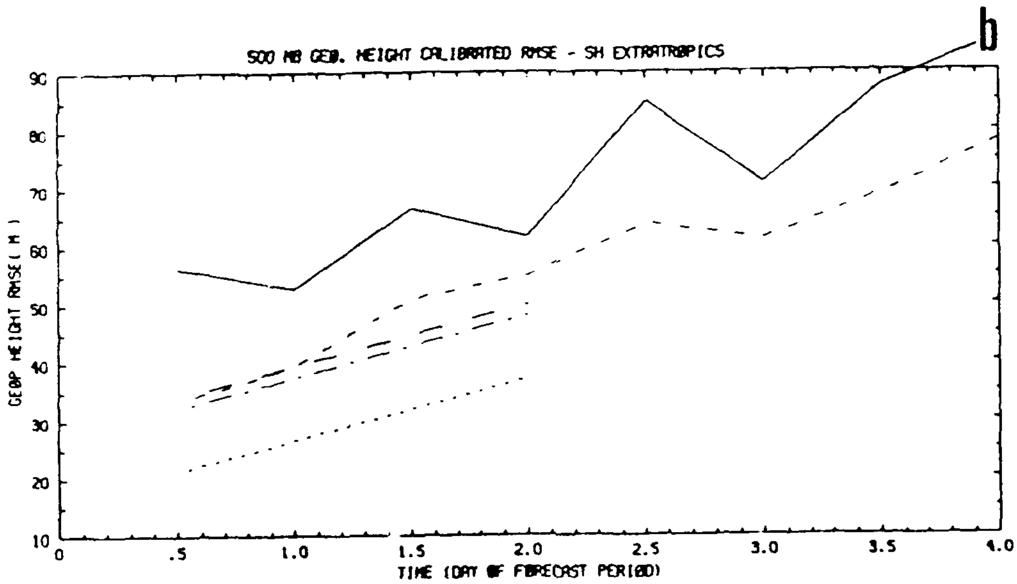


Fig. 20 Calibrated 500 mb height rms errors for Southern Hemisphere extratropics. *(a) Winter OSEs, (b) Summer OSE's.*

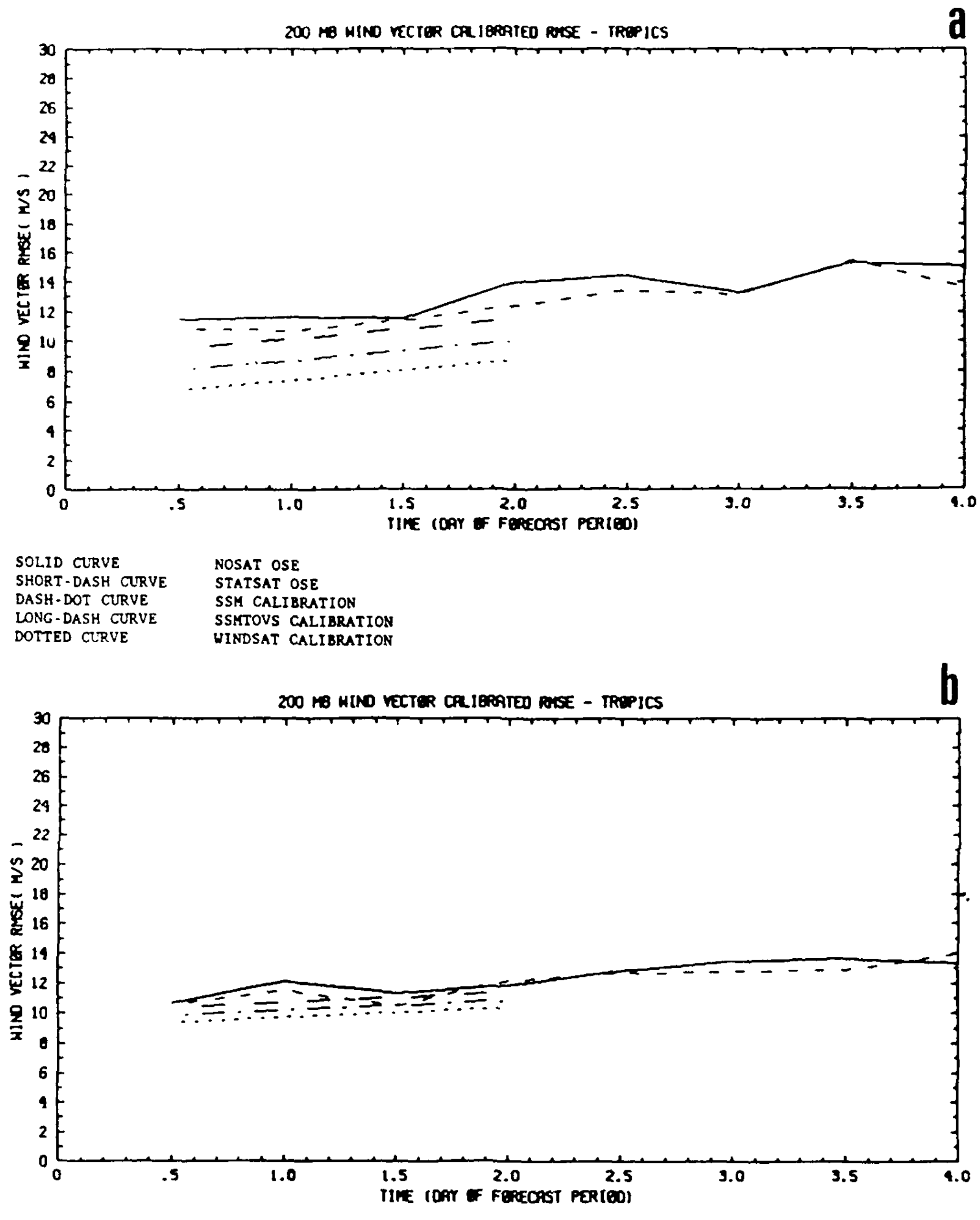


Fig. 21 Calibrated 200 mb wind vector rms errors for tropics. (a) Winter OSEs. (b) Summer OSEs.

differences are so small, it is difficult to judge the impacts by eye. In this case, in terms of predictability time there are significant impacts. WINDSAT has a 24 hour advantage over SSM, a 48 hour advantage over SSM+TOVS and a 60 hour advantage over STATSAT. Further SSM+TOVS has approximately a 12 hour advantage over STATSAT which in turn has a 12 hour advantage over NOSAT. However, not much weight should be given these results since in the tropics and Southern Hemisphere the calibration procedure is not very reliable.

For the Northern Hemisphere 850 mb relative humidity or cloud cover (Fig. 22), the actual impacts are all expected to be rather small. Note again the close relationship between relative humidity and cloud cover statistics.

The calibration procedure does have some uncertainties and drawbacks. The main drawback is the assumption of a linear relationship between impact in the OSSEs and in the OSEs. Of course the data assimilation system and nature are highly nonlinear. In the current experiments, the uncertainties are mostly due to the small sample size, especially in the Southern Hemisphere, where the number of radiosondes used in the verification is small. For example, in some cases the sense of impact between NOSAT and STATSAT is reversed in OSSEs and OSEs. In these cases the calibration produces nonsensical results. This occured when calibrating the Southern Hemisphere 850 mb relative humidity rms differences. In other cases the OSE impact between NOSAT and STATSAT is quite small. This implies negligible calibrated impact for any change to the data assimilatian system. This occurred when calibrating the Northern Hemisphere 500 mb height rms differences.

5.4 Forecast Biases

In general the biases during the forecast are small compared to the rms differences. However in many cases the biases grow very steadily with time indicating that the AFGL model is warming and drying relative to the ECMWF nature.

We examined the biases by fitting straight lines with a common slope as described in Section 5.2. However in this case all data from 12 to 96

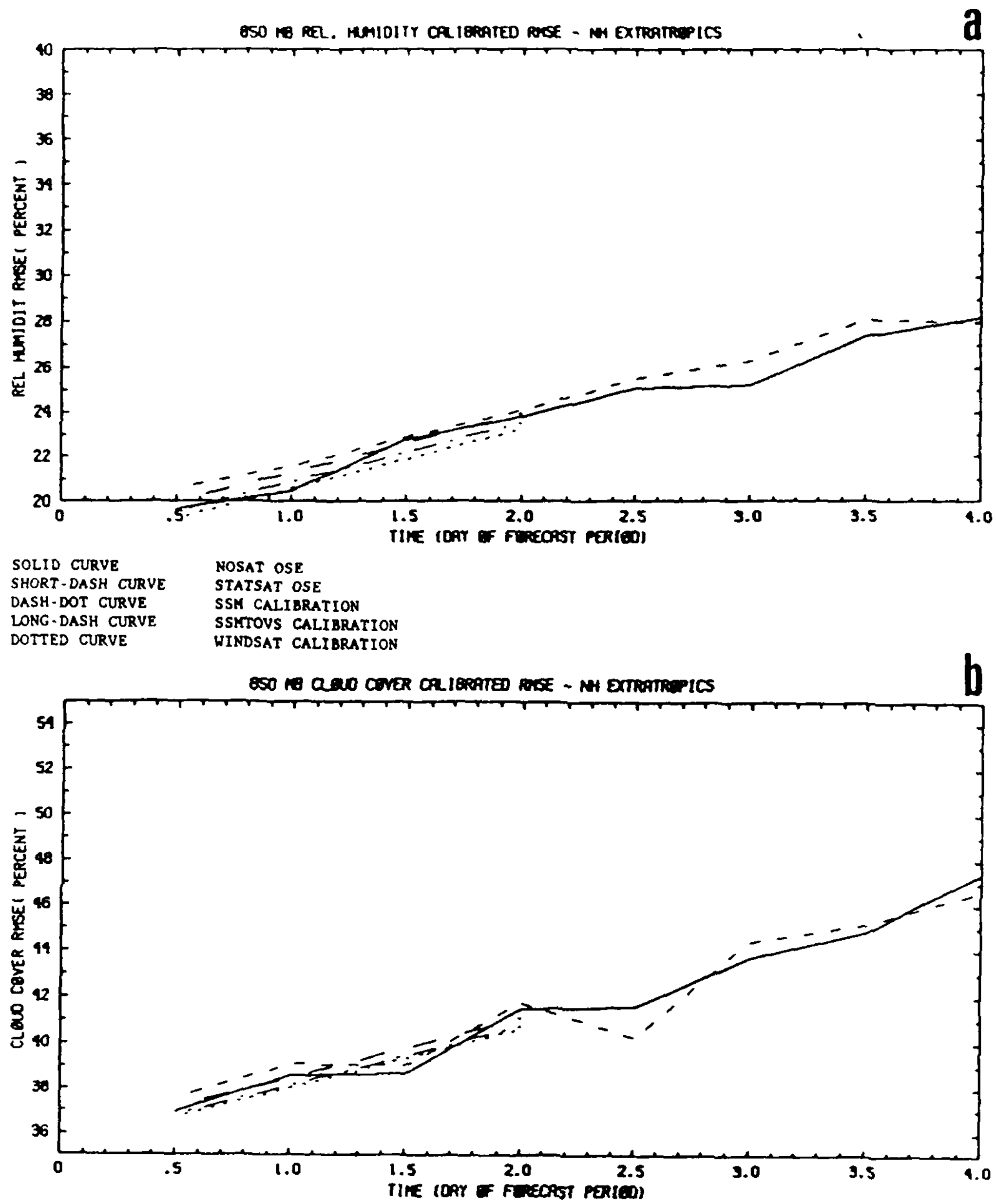


Fig. 22 Calibrated 850 mb summer OSE rms errors for Northern Hemisphere extratropics. (a) Relative humidity, (b) Cloud cover.

Table 6. Growth rate of forecast bias for height (m/day), temperature (K/day) and relative humidity (%/day) for the N.H. extratropics. Impacts for cases with R**2 less than 0.64 (i.e. for correlations less than 0.8) are not shown.

l	P(l)	R**2	Z	R**2	T	R**2	RH
1	1000	.000	-	.175	-	.673	-1.467
2	850	.561	-	.879	.443	.753	-.948
3	700	.963	4.232	.983	.718	.848	-1.074
4	500	.989	12.625	.991	.937	.284	-
5	400	.992	18.747	.986	.903	.014	-
6	300	.992	25.569	.986	.689	.057	-
7	250	.993	28.995	.985	.544	-	-
8	200	.994	31.843	.979	.426	-	-
9	150	.995	35.529	.955	.371	-	-
10	100	.996	39.334	.970	.375	-	-
11	70	.998	43.682	.964	.406	-	-
12	50	.997	47.277	.973	.415	-	-

hours was used. The common slope obtained from the fit is then the rate at which the biases grow. Some of these are displayed in Table 6 for the Northern Hemisphere OSSEs. Generally speaking the forecasts are warming relative to nature by one third to one degree per day. The height biases reflect these temperature biases. These results are consistent with the warming seen during the spinup forecast (Section 4.1) Also the forecasts are drying in the lower atmosphere by 1 to 1.5 percentage points of relative humidity per day. For temperature and height there are many cases when the fraction of variance explained by the fit is greater than .99, indicating that the bias grows very linearly. For example Fig. 23 shows the evolution of bias for the 500 mb height in the OSSEs. Differences between the different experiments are not significant. Results for the tropics and Southern Hemisphere are not as regular and clear cut presumably because of sampling variability. This also applies to the OSEs, although there is some evidence of the forecasts warming during the February OSEs in the mid troposphere. In the Northern Hemisphere, the dry relative humidity biases are substantial at 1000 mb. In this case, the bias at the start of the forecast is already -17 to -18 percent. At other levels, the initial dry bias is only of order 5%. Typically, the analyses are dry by 5% and the forecasts continue to dry out by 1% to 2% per day for the first two days of the forecast. For example, the evolution of the biases of the 850 mb relative humidity forecasts are shown in Fig. 24. Again, differences between the different experiments are not significant and the trends are not so clear cut in the tropics and extratropics and the OSEs. In the tropics, in the OSSEs, the atmosphere tends to moisten during the forecast. In this case the boundary layer is analyzed dry but the mid troposphere is slightly moist.

6. Summary and Conclusions

We have conducted a series of state of the art observing system simulation experiments (OSSEs) to assess the impact of a Doppler Wind Lidar (DWL) sounder. The addition of DWL profiles in our WINDSAT experiment significantly improved the initial state specification, especially in the Southern Hemisphere extratropics relative to our control STATSAT

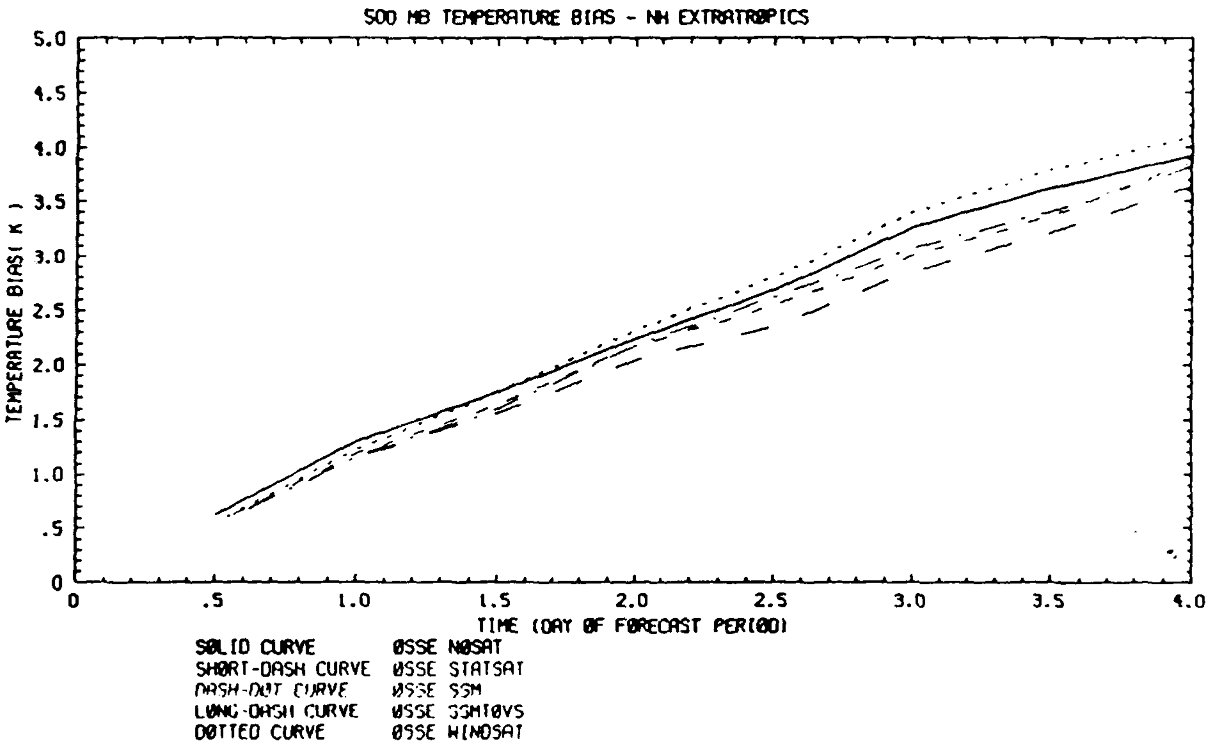


Fig. 23 Forecast 500 mb temperature bias for Northern Hemisphere extratropics.

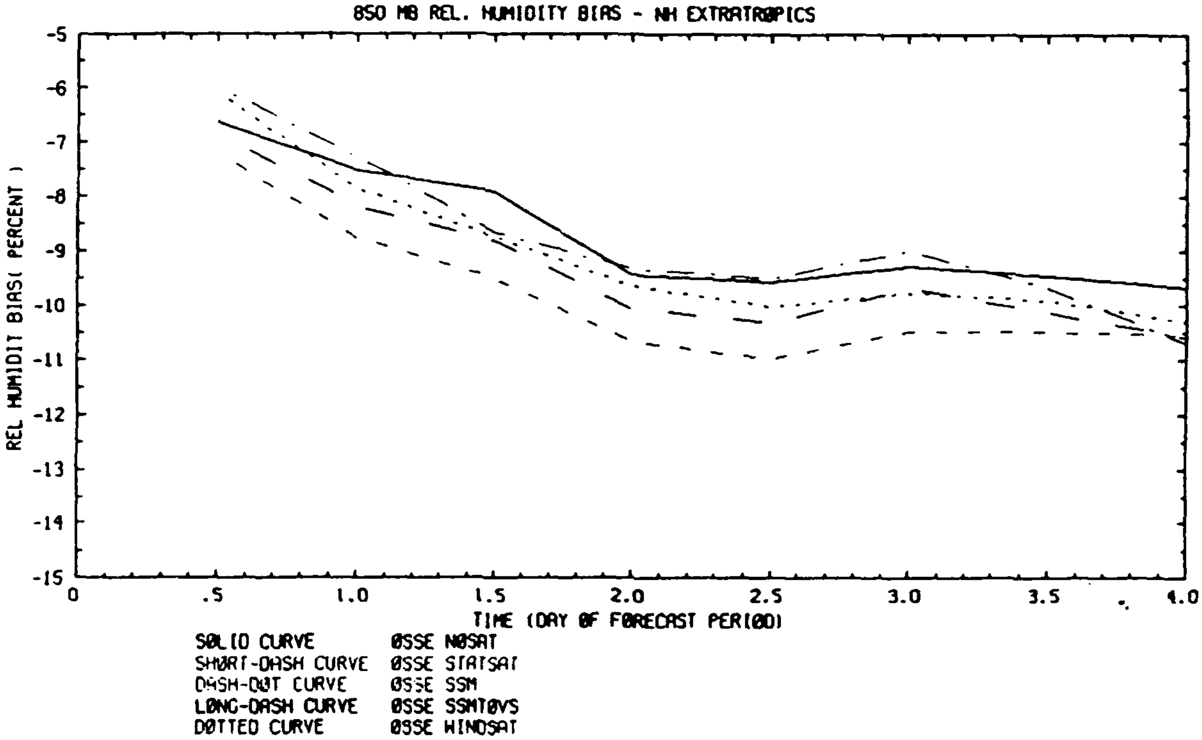


Fig. 24 Forecast 850 mb relative humidity bias for Northern Hemisphere extratropics.

experiment. In order to infer realistic impacts our OSSE results are calibrated against results from (real data) observing system experiments (OSEs) described in more detail by Louis et al. (1988). Our principal findings are the listed below, but first we note some caveats. First the error characteristics chosen for the DWL are optimistic, both in terms of their small magnitude and randomness. Second our calibrated results rely on a number of assumptions and the size of the Southern Hemisphere and tropical radiosonde sample limit their reliability.

1. WINDSAT analyses and forecasts are much better than any of the other analyses and forecasts. The WINDSAT data coverage and quality is good; furthermore the WINDSAT errors are uncorrelated, although the data is dense. Improvements in forecasting ability were quite large, in the Southern Hemisphere. These differences are expected to increase the length of the useful forecast by 36 hours in the height field at 500 mb and by 48 hours in the wind field at 200 mb.

2. The WINDSAT impacts in extratropics are very significant. We suspect that the use of full multivariate OI may be necessary to gain full usefulness from DWL data. In particular, according to adjustment theory (Blumen, 1972) the large scale extratropical wind field should adjust to the mass field. Therefore it is important to balance the analysis increments due to the WINDSAT data with corresponding mass field increments. Otherwise the extratropical WINDSAT data will tend to be rejected.

3. Details of the analyzed tropical wind field were somewhat disappointing. Although the size of the impacts in terms of rms vector wind errors was small, there were definite improvements in the predictability of the tropical winds. However the mean meridional circulation, as evidenced by $[\bar{v}]$ and the tropical divergence were not especially better analyzed with the WINDSAT data. Improvements to the assimilation procedures might enable the WINDSAT data to have greater impacts in the tropics. Currently, the wind analysis increments are required to be in geostrophic balance by the analysis

procedure. Furthermore the initialization procedure does not include effects of convection.

4. Improved wind data also improved the analyzed and forecast moisture and cloudiness fields. Relative humidity forecasts are best in WINDSAT although SSM had better relative humidity analyses. This is to be expected since the relative humidity field adjusts to the large scale mass-motion fields, which are better analyzed and forecast in WINDSAT.

5. The AFGL model has a tendency to warm and dry out relative to the ECMWF nature model. This warming is seen in all the forecasts. We note that the version of the AFGL model which we used has no radiation parameterization and hence no cooling mechanism although there is a constant source of warming due to the fixed sea surface temperature.

6. Cloud cover estimates derived from the relative humidity fields are too high. Either the model is too moist or the relative humidity to cloud cover algorithm needs to be tuned. In any case, cloud cover differences or comparisons are still useful since all relative humidity fields converted to cloud cover will be too cloudy in the same way.

7. The comparisons of rms differences of cloudiness yield the same results as comparisons of rms differences of relative humidity. This might have been expected in view of the facts that the statistics calculated are averages over large samples and that the relative humidity to cloud cover relationship is simple. This relationship is nonlinear but it is 1-to-1 and monotonic and it does not depend on any other model parameters known to impact cloudiness such as temperature lapse rates, vertical wind shears, vertical velocity or divergence.

8. Impacts in the Northern Hemisphere forecasts were larger when calculated as grid point rms errors than when calculated as radiosonde rms differences. As seen in the synoptic charts the greatest impacts tend to be polar, however the verifying radiosondes tend to be midlatitude.

9. The calibration indicates that the improvements seen in the OSSEs in the Southern Hemisphere and tropics are realistic, but in the Northern Hemisphere extratropics, the fact that satellite data has little impact as seen in our NOSAT versus STATSAT comparisons implies that any novel observing system will have limited impact.

10. Our results are generally consistent, yet different in details with previous studies. Atlas et al. found 24 hour improvements in the Southern Hemisphere while we obtained 36 hour improvements (for 500 mb heights) and more for the upper level winds. Atlas et al. used a univariate analysis system and had much more skillful NOSAT and STATSAT (their Control and FGGE) analyses and forecasts.

There is considerable opportunity to improve and refine the experiments reported here and elsewhere. Such efforts would allow the quantification of the relative impact of proposed advanced temperature sounders and DWLs. In addition cost benefit analyses of observational accuracies could be supported by such studies. In future studies it will be important to carefully simulate the geographical coverage and error characteristics of proposed instruments. In particular, natural phenomena which give rise to correlated observational errors should be included to the extent possible. We mention two such phenomena in the following paragraphs. It is also important that the assimilation system be modified to best take advantage of the novel observations; for example for DLWs or other tropical wind observing systems the analysis procedure for the tropical winds and the normal mode initialization should be improved to allow for divergent wind increments and to include the effects of convection.

The error characteristics and distribution of simulated lidar winds for example should depend on the global distribution of aerosols and clouds. These geophysical parameters are in turn associated (correlated) in the real world with the geophysical parameters which are to be measured. This is one cause of spatial and temporal error correlations and as such it should be included in our simulation of observational errors. Hence, it is important that the model used to generate nature also can provide a realistic description of the aerosol and cloud fields, as mentioned in the third point above. In future work we anticipate using a newer nature run generated by the ECMWF using a T106 spectral truncation and more complete physical parameterizations. This nature run includes many diagnostic fields generated by the physical parameterizations.

No existing global model has fine enough resolution to represent all scales of motion which exist in nature. In fact the smallest scales represented by models are usually severely damped for computational reasons. In real data assimilation these small scales are considered part of the observational error. In fact for radiosondes this is the greatest source of error. Although the length scale is small, these errors are correlated. Therefore simulated observations should include spatially correlated errors. This could be accomplished by unfiltering the nature run, thereby restoring the smallest resolvable scales as suggested by Hoffman (1988).

Acronyms

AFGL	Air Force Geophysics Laboratory
ASAP	AFGL Statistical Analysis Program
CDW	cloud drift wind
COBAL	COnstant level BALloon
DWL	Doppler wind lidar
ECMWF	European Centre for Medium-Range Weather Forecasts
FGGE	First GARP Global Experiment
GFDL	Geophysical Fluid Dynamics Laboratory
GLA	GFSC Laboratory for Atmospheres (NASA)
GSM	global spectral model
LIMS	stratospheric sounding data
NESDIS	National Environmental Satellite and Data Information Service
NMC	National Meteorological Center
NMI	normal mode initialization
NOSAT	OSE or OSSE using NO SATellite data
NWP	numerical weather prediction
OESD	observing error standard deviation
OI	optimal interpolation
OSE	observing system experiment
OSSE	observing systems simulation experiment
RAOB	radiosonde observation
RH	relative humidity
SPINUP	SPIN UP OSE or OSSE
SSM	special sensor microwave
STATSAT	OSE or OSSE using conventional satellite data
TOVS	TIROS operational vertical sounder
WINDSAT	OSSE using simulated DWL data
WMO	World Meteorological Organization

7. References

Arnold, C.P., C.H. Dey and W.J. Bostelman, 1985: Results of an observing system simulation experiment based on the proposed WINDSAT instrument. In Global Wind Measurements, edited by W.E. Baker and R.J. Curran, A. Deepak Publishing, Hampton, VA, pp. 81-83.

Atlas, R., E. Kalnay, W.E. Baker, J. Susskind, D. Reuter, and M. Halem, 1985: Simulation studies of the impact of future observing systems on weather prediction. In Preprint volume. Seventh Conference on Numerical Weather Prediction. American Meteorological Society, Boston, MA, pp. 145-151.

Baker, W.E., and R.J. Curran, (Eds.), 1985: Report of the NASA Workshop on Global Wind Measurements, A. Deepak Publishing, Hampton, VA, 50 pp.

Ballish, B., 1980: Initialization, theory and application to the NMC spectral model. Ph.D. Thesis, Dept. of Meteorology, University of Maryland, College Park, MD, 151 pp.

Bengtsson, L., M. Kanamitsu, P. Kållberg, and S. Uppala, 1982: FGGE 4-dimensional data assimilation at ECMWF. Bull. Am. Meteorol. Soc., 63, 29-43.

Bergman, K. H., 1979: Multivariate analysis of temperature and winds using optimum interpolation. Mon. Weather Rev., 107, 1423-1444.

Blumen, W., 1972: Geostophic adjustment. Rev. Geophys. Space Phys., 10, 485-528.

Brenner, S., C.-H. Yang, and S.Y.K. Yee, 1982: The AFGL spectral model of the moist global atmosphere: Documentation of the baseline version. AFGL-TR-82-0393, AFGL, Hanscom AFB, MA, 65 pp. [NTIS AD A129283]

Brenner, S., C.-H. Yang, and K. Mitchell, 1984: The AFGL global spectral model: Expanded resolution baseline version. AFGL-TR-84-0308. [NTIS ADA160370]

Curran, R.J., et al., 1988: LAWS Instrument Panel Report, NASA, Washington, DC, 55 pp.

Dalcher, A., and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. Tellus, 39A, 474-491.

Daley, R., and T. Mayer, 1986: Esimates of global analysis error from the global weather experiment observational network. Mon. Weather Rev., 114, 1642-1653.

Dey, C.H., and L.L. Morone, 1985: Evolution of the National Meteorological Center Global Data Assimilation System: January 1982-December 1985. Mon. Weather Rev., 113, 304-318.

Dey, C., C.P. Arnold, and W. Bostelman, 1985: Design of a windsat observing system simulation experiment. In Global Wind Measurements, W. E. Baker and R. J. Curran (Eds.), A. Deepak Publishing, Hampton, VA, pp 73-79.

Fye, F.K., 1978: AFGWC automated cloud analysis model. Report AFGWC-TM-78-007, 97 pp.

Halem, M., E. Kalnay, W.E. Baker, and R. Atlas, 1982: An assessment of the FGGE satellite observing system during SOP-1. Bull. Am. Meteorol. Soc., 63, 407-426.

Hoffman, R., 1988: Sampling data for OSSEs. Preprints, Eighth Conference on Numerical Weather Prediction, February 22-26, Baltimore, MD, Amer. Meteor. Soc.

Hollingsworth, A., K. Arpe, M. Tiedtke, M. Capaldo and H. Savijärvi, 1980: The performance of a medium range forecast model in winter - impact of physical parameterization. Mon. Weather Rev., 108, 1736-1773.

Lau, N.-C., 1984: Circulation statistics based on FGGE Level III-b analyses. NOAA Data Report ERL GFDL-5, NOAA, Washington DC, 427 pp.

Louis, J.-F., R.N. Hoffman, T. Nehrkorn, T. Baldwin, and M. Mickelson, 1988: Four-dimensional data assimilation experiments using the AFGL GDAS. Preprints, Eighth Conference on Numerical Weather Prediction, February 22-26, 1988, Baltimore, MD., AMS, Boston, MA.

McPherson, R.D., K.H. Bergman, R.E. Kistler, G.E. Rasch, and D.S. Gordon, 1979: The NMC operational global data assimilation system. Mon. Weather Rev., 107, 1445-1461.

Norquist, D.C., 1986: Alternative forms of moisture information in 4-D data assimilation. AFGL-TR-86-0194. [NTIS ADA179792]

Norquist, D.C., 1988: Alternate forms of humidity information in global data assimilation. Mon. Wea. Rev., 116, 452-471.

Oort, A.H., 1983: Global atmospheric circulation statistics, NOAA Professional Paper 14, U.S. Department of Commerce, Washington, DC., 180 pp. [NTIS PB8-4-129717].

Salvetti, G., 1987: Spaceborne doppler wind lidars. ESA Journal 1987, 11, 19.

Schlatter, T., 1981: An assessment of operational TIROS-N temperature retrievals over the United States. Mon. Weather Rev., 109, 110-119.

Sela, J.G., 1980: Spectral modeling at the National Meteorological Center. Mon. Weather Rev., 108, 1279-1292.

Tiedtke, M., J.-F. Geleyn, A. Hollingsworth and J.-F. Louis, 1979: ECMWF model parameterization of subgrid scale processes. <u>ECMWF Tech. Rep. No. 10.</u> ECMWF, Reading, Berkshire, U.K.

WMO, 1986: Formats for International Exchange of FGGE Level II Data Sets, International Council of Scientific Unions/World Meteorological Organization, The World Climate Research Programme, <u>WMO/TD-No. 100</u>, WMO, Geneva, 86 pp.